Université de Montréal

Décoder les émotions à travers la musique et la voix

par Sébastien Paquette

Département de Psychologie
Faculté des Arts et des Sciences

Thèse présentée
en vue de l'obtention du grade de doctorat
en Psychologie
option Sciences Cognitives et Neuropsychologie

Décembre, 2016

Résumé

L'objectif de cette thèse est de comparer les mécanismes fondamentaux liés à la perception émotionnelle vocale et musicale. Cet objectif est sustenté par de nombreux rapports et théories appuyant l'idée de substrats neuronaux communs pour le traitement des émotions vocales et musicales. Il est proposé que la musique, afin de nous faire percevoir des émotions, recrute/recycle les circuits émotionnels qui ont évolué principalement pour le traitement des vocalisations biologiquement importantes (p.ex. cris pleurs). Bien que certaines études ont relevé de grandes similarités entre ces deux timbres (voix, musique) du point de vue cérébral (traitement émotionnel) et acoustique (expressions émotionnelles), certaines différences acoustiques et neuronales spécifique à chaque timbre ont également été observées. Il est possible que les différences rapportées ne soient pas spécifiques au timbre, mais observées en raison de facteurs spécifiques aux stimuli utilisés tels que leur complexité et leur longueur. Ici, il est proposé de contourner les problèmes de comparabilité de stimulus, par l'utilisation des expressions émotionnelles les plus simples dans les deux domaines.

Pour atteindre l'objectif global de la thèse, les travaux ont été réalisés en deux temps. Premièrement, une batterie de stimuli émotionnels musicaux comparables aux stimuli vocaux déjà disponibles (Voix Affectives Montréalaises) a été développée. Des stimuli (Éclats Émotionnels Musicaux) exprimant 4 émotions (joie, peur, tristesse, neutralité) performés au violon et à la clarinette ont été enregistrés et validés. Ces Éclats Émotionnels Musicaux ont obtenu un haut taux de reconnaissance (M=80.4%) et reçu des jugements d'arousal (éveil/stimulation) et de valence correspondant à l'émotion qu'il représentait. Nous avons donc pu, dans un deuxième temps, utiliser ces stimuli nouvellement validés et les Voix

Affectives Montréalaises pour réaliser deux études de comparaison expérimentales. D'abord, nous avons effectué à l'aide de l'imagerie par résonnance magnétique fonctionnelle une comparaison des circuits neuronaux utilisés pour le traitement de ces deux types d'expressions émotionnelles. Indépendamment de leur nature vocale ou musicale, une activité cérébrale spécifique à l'émotion a été observée dans le cortex auditif (centrée sur le gyrus temporal supérieur) et dans les régions limbiques (gyrus parahippocampique/amygdale), alors qu'aucune activité spécifique aux stimuli vocaux ou musicaux n'a été observée. Par la suite, nous avons comparé la perception des émotions vocales et musicales sous simulation d'implant cochléaire. Cette simulation affectant grandement la perception des indices acoustiques liés aux hauteurs tonales (important pour la discrimination émotionnelle), nous a permis de déterminer quels indices acoustiques secondaires à ceux-ci sont importants pour la perception émotionnelle chez les utilisateurs d'implant cochléaire. L'examen des caractéristiques acoustiques et des jugements émotionnels a permis de déterminer que certaines caractéristiques timbrales (clarté, énergie et rugosité) communes à la voix et la musique sont utilisées pour réaliser des jugements émotionnels sous simulations d'implant cochléaire, dans les deux domaines.

L'attention que nous avons portée au choix des stimuli nous a permis de mettre de l'avant les grandes similarités (acoustique, neuronales) impliquées dans la perception des émotions vocales et musicales. Cette convergence d'évidence donne un appui important à l'hypothèse de circuits neuronaux fondamentaux commun pour le traitement des émotions vocales et musicales.

Mots-clés : Musique, Voix, Émotions, Comparaisons Inter-domaines, Caractéristiques Acoustiques, Timbre, IRMf, Implant Cochléaire

Abstract

The aim of this thesis is to compare the fundamental mechanisms related to vocal and musical emotion perception. This objective is supported by many reports and theories bringing forward the idea of common neural substrates for the treatment of vocal and musical emotions. It is proposed that music, in order to make us perceive emotions, recruits/recycles the emotional circuits that evolved mainly for the treatment of biologically important vocalisations (e.g. cries, screams). Although some studies have found great similarities between these two timbres (voice, music) from the cerebral (emotional treatment) and acoustic (emotional expressions) point of view, some acoustic and neural differences specific to each timbre have also been reported. It is possible that the differences described are not specific to the timbre but are observed due to factors specific to the stimuli used such as their complexity and length. Here, it is proposed to circumvent the problems of stimulus comparability by using the simplest emotional expressions in both domains.

To achieve the overall objective of the thesis, the work was carried out in two stages. First, a battery of musical emotional stimuli comparable to the vocal stimuli already available (Montreal Affective Voices) was developed. Stimuli (Musical Emotional Bursts) expressing 4 emotions (happiness, fear, sadness, neutrality) performed on the violin and clarinet were recorded and validated. These Musical Emotional Bursts obtained a high recognition rate (M = 80.4%) and received arousal and valence judgments corresponding to the emotion they represented. Secondly, we were able to use these newly validated stimuli and the Montreal Affective Voices to perform two experimental comparison studies. First, functional magnetic resonance imaging was used to compare the neural circuits used to process these two types of emotional expressions. Independently of their vocal or musical nature, emotion-specific activity was observed in the auditory cortex (centered on the superior temporal gyrus) and in the limbic regions (amygdala/parahippocampal gyrus), whereas no activity specific to vocal or musical stimuli was observed. Subsequently, we compared the perception of vocal and musical emotions under cochlear implant simulation. This simulation greatly affects the perception of acoustic indices related to pitch (important for emotional discrimination), allowing us to determine which acoustic indices secondary to these are important for emotional perception in cochlear implant users. Examination of acoustic characteristics and emotional judgments determined that certain timbral characteristics (brightness, energy, and roughness) common to voice and music are used to make emotional judgments in both domains, under cochlear implant simulations.

The specific attention to our stimuli selection has allowed us to put forward the similarities (acoustic, neuronal) involved in the perception of vocal and musical emotions. This convergence of evidence provides important support to the hypothesis of a fundamental common neural circuit for the processing of vocal and musical emotions.

Keywords : Music, Voice, Emotion, Cross-domain Comparison, Acoustic Features, Timbre, fMRI, Cochlear Implant

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

Francophones:

COF: Cortex Orbitofrontal

GFI: Gyrus Frontal Inférieur

GTS: Gyrus Temporal Supérieur

IC: Implant Cochléaire

IRMf: Imagerie par Résonnance Magnétique fonctionelle

MAV: Voix Affectives Montréalaises

MEB: Éclats Émotionnels Musicaux

MVPA: Analyse de Patrons Multivoxels

ROI: Région d'Intérêt

STS: Sulcus Temporal Supérieur

Anglophones:

AMYG: Amygdala

ANOVA: Analysis of variance

AC: Auditory Cortex

BOLD: Blood-oxygen-level dependent

CI: Cochlear Implant

EPI: Echo Planar Imaging

fMRI: functional Magnetic Resonance Imaging

FWE: Family-Wise Error

GLM: General Linear Model

HG: Heschl's Gyrus

Hz: Hertz

M: Mean

MM: Metronome Mark

MAV: Montreal Affective Voices

MBEA: Montreal Battery of Evaluation of Amusia

MEB: Musical Emotional Burst

MIDI: Musical Instrument Digital Interface

PAID: parallel-acquired inhomogeneity-desensitized

PHG: Parahippocampal Gyrus

PT: Planum temporale

RMS: Root Mean Square

SD: Standard Deviation

SE: Standard Error

SIC: Spectral Information Change

SPM: Statistical Parametric Mapping

STG: Superior Temporal Gyrus

TE: Echo Time

TR: Repetition Time

‘‘Music is the universal language of Mankind’’

– Henry Wadsworth Longfellow (1857)

Remerciements

Durant cette thèse j'ai eu la chance de travailler avec deux brillants superviseurs Isabelle Peretz et Pascal Belin.

Merci, Isabelle, d'avoir été une mentore idéale ces 10 dernières années. Vous avez su balancer mes avidités en me permettant d'explorer tous les projets qui me venaient en tête, même s'ils me menaient à l'étranger. Votre support constant, votre compréhension et votre rigueur sont la pierre angulaire de cette thèse.

Merci, Pascal, pour votre ténacité et générosité académique qui m'ont permis de dépasser mes limites. Les apprentissages que j'ai réalisés sous votre supervision vont me permettre d'être bien équipé pour affronter mes prochains objectifs.

Merci à mes collègues du Brams, du Cernec et à mes amis pour votre support constant. Merci d'avoir supporté mes absences comme mes excès. Merci à Isabelle, Geneviève, Patrice et Kathrin, nos interactions quotidiennes et vos encouragements m'ont permis de rester saint d'esprit ou du moins aussi saint d'esprit que vous. Vincent, Sara et Louis votre rigueur scientifique en conférence et après 5h a permis de faciliter tous mes apprentissages. Véronique, Synthia, Marie-Élaine, Alex, Jen, Dahlia, Sarah, Mona, Dominique, Philippe, Mihaela, Jocelyne, Bianca, Pauline, Benjamin, Bruno, Flor, David, Alexandre, Nathalie, Séverine et Bernard vous avez contribué à cette thèse plus que vous le pensez.

Merci à mes parents et à mon frère, votre support dans l'atteinte de mes objectifs a débuté bien avant cette thèse et ma reconnaissance envers vous pour celui-ci est inquantifiable. À vous, merci pour tout.

Chapitre 1: Contexte Théorique

1.1 La communication émotionnelle

La communication des émotions, les percevoir et les produire, a été décrite par Ekman (1992) comme cruciale aux relations sociales et à la survie. L'importance de cette communication ne se limite pas qu'aux êtres humains, les primates non humains ont un riche répertoire d'expressions vocales leur permettant de transmettre des signaux précis aux membres de leur groupe social (p.ex. Winter, Ploog, & Latta, 1966), que ce soit pour les avertir d'un danger ou pour les mobiliser.

Chez l'humain, les émotions communiquées de manière auditive ont une importance et une utilité similaire. Plus particulièrement, il a été démontré que les expressions vocales non linguistiques détiennent certains avantages lors d'interactions sociales. Comme chez les primates elles permettent de rejoindre un grand nombre d'individus dispersés sur un large territoire. Leur efficacité pour attirer l'attention des autres vient du fait que les émotions perçues de manière auditive ne requièrent pas de contact (visuel) direct pour être communiquées (Hawk, van Kleef, Fischer, & van der Schalk, 2009). Cette grande efficacité rend la perception d'émotions auditives fondamentale pour nos interactions sociales, car elle affecte et structure nos comportements; elle nous permet d'interagir et de vivre en société. Effectivement, bien comprendre les signaux auditifs de notre environnement (p.ex. cris, pleurs) nous permet de continuellement nous adapter selon la situation à laquelle nous faisons face. Ce qui rend l'étude des processus liés à la perception émotionnelle auditive d'un grand intérêt, tant au niveau fondamental (biologique/neurologique) que chez des populations ayant des difficultés à les percevoir.

L’habileté de l’être humain à communiquer les signaux émotionnels s’est raffinée à travers nos diverses sociétés, et permet maintenant de communiquer de nombreuses émotions. Certains modèles (p.ex. Yik, Russell, & Barrett, 1999) proposent que l’ensemble des émotions varient sur deux dimensions affectives, leur niveau d’excitabilité/éveil ‘‘Arousal’’ (de peu stimulant à très stimulant) et leur valence (de négatif à positif). Par exemple, certaines émotions sont perçues comme ayant un très haut niveau d’excitabilité (p.ex. enthousiasme, nervosité) et d’autres peu (p.ex. déception, satisfaction). Orthogonalement, ces mêmes émotions peuvent être différenciées par leur valence; les expressions d’enthousiasme et de satisfaction sont perçues comme étant plus positives tandis que celles exprimant la nervosité et la déception sont perçues comme plus négatives.

Malgré ce grand univers émotionnel, certaines émotions dites de base (Ekman, 1992) se sont développées pour leur valeur adaptative permettant de répondre à des tâches fondamentales de la vie courante. Ces émotions permettraient de signifier, par exemple, l’atteinte d’objectifs, l'évaluation des pertes ou de la frustration (Johnson-Laird & Oatley, 1992); des situations qui font partie des thèmes centraux relationnels. Pour être considérées comme une émotion de base, ces émotions doivent être universelles, avoir des expressions distinctes (perçues rapidement) et être reconnues dès un très jeune âge (innées). Selon Ekman (1992), qui a majoritairement étudié les expressions faciales, six expressions répondent à ces critères : la joie, la peur, la colère, la tristesse, le dégout et la surprise.

Depuis ces travaux emblématiques, la majorité des études sur les émotions ont été réalisés avec des stimuli visuels (visages, postures). De même que les émotions communiquées visuellement, les émotions communiquées de manière auditive (vocales et musicales) sont également reconnues dès un très jeune âge, de manière universelle et de façon distincte.

1.1.1 Les émotions vocales – le visage auditif

Nous sommes des experts de la voix; nous avons probablement passé plus de temps quotidiennement, depuis notre jeune âge, à écouter des voix que tout autres sons. Celle-ci nous permet de communiquer, de nous organiser, et de transmettre des informations importantes comme notre identité et notre état affectif (Belin, Fecteau, & Bédard, 2004).

Des études en neuroimagerie ont permis d'établir que la perception de ces émotions vocales se développe très tôt ou serait innée. Une étude en particulier (Grossmann, Oberecker, Koch, & Friederici, 2010) a démontré chez l'enfant que dès le septième mois de vie l'on pouvait moduler, par la prosodie[1] émotionnelle, l'activité cérébrale dans le cortex supérieur temporal droit. Chez une population similaire, des enfants de 3 à 7 mois, Blasi et collaborateurs (2011) ont également observé que les vocalisations de tristesse modulaient l'activité dans le cortex orbito-frontal et l'insula. Ces résultats suggèrent que très tôt dans notre développent (dès la petite enfance) nous possédons des circuits neuronaux permettant de discriminer les émotions vocales de notre propre culture.

Il a également été observé que nous sommes en mesure de déduire les émotions vocales exprimées dans une culture à laquelle nous n'avons jamais été exposés. Cette capacité des émotions vocales à traverser les cultures, supportant l'universalité de celles-ci, a été observée chez des Occidentaux (Canadiens) parlant anglais (Thompson & Balkwill, 2006).

---

[1] Ensemble de paramètres acoustiques du langage qui permet à l'auditeur de déduire une grande partie de l'état affectif du locuteur (Scherer, 1995).

Lors d'une tâche où ils devaient juger l'intention émotive de locuteurs de différentes langues (anglais, allemand, chinois, japonais, et tagalog) qui lisaient des énoncés (au contenu verbal neutre) avec une prosodie évoquant une émotion (joie, colère, tristesse et peur), leur identification des émotions s'est révélée être au-dessus du hasard pour toutes les émotions dans toutes les langues. Plusieurs études comme celle-ci ont démontré le caractère universel de la prosodie émotionnelle (p.ex. Pell, Monetta, Paulmann, & Kotz, 2009, pour de résultats similaires avec des participants Argentins).

Non seulement la prosodie émotionnelle traverse les cultures, mais l'analyse des paramètres acoustiques qui la compose se réalise rapidement; en moyenne les adultes ont besoin d'une demi-seconde à une seconde, pour reconnaitre des émotions vocales comme la peur, la joie, la tristesse et la neutralité (Pell & Kotz, 2011).

Comme précisé précédemment, il n'y a pas que la prosodie qui permet de capter cet état affectif; nous sommes dotés de capacités à extraire des informations paralinguistiques dans la voix; nous pouvons interpréter une simple vocalisation (un cri, un rire) sans qu'elle contienne des mots (Scherer, 1986). Ces éclats émotionnels non verbaux (Scherer, 1994; également appelés interjections non verbales, Schröder, 2003) sont généralement associés à un état émotionnel intense. Ils sont décrits comme reflétant plus une poussée biologique que sociologique (Scherer, 1986); ils sont plus similaires aux expressions primitives des bébés et des animaux que de la parole émotionnelle.

Ces éclats émotionnels non verbaux sont peu stylisés, donc représentent un moyen relativement universel de la communication humaine spontanée. Sauter, Eisner, Ekman, & Scott (2010; 2015) ont évalué la reconnaissance des vocalisations émotionnelles non verbales, auprès de deux groupes culturels radicalement différents; des Occidentaux du Royaume-Uni et des individus d’un village nambien isolé. Les auteurs ont observé que les vocalisations émotionnelles exprimant la colère, le dégout, la peur, la joie, la tristesse et la surprise pouvaient être reconnues au-dessus du hasard par les deux groupes. Des résultats similaires supportant le caractère universel de ces expressions spontanées ont été obtenus lors de comparaisons d’autres groupes culturels relativement différents: des Canadiens et des Japonais (Koeda, Belin, Hama, Masuda, & Matsuura, 2013), des Équatoriens et des Américains (Bryant & Barrett, 2008) et des individus provenant de neuf pays d’Europe, d’Amérique et d’Asie (Scherer, Banse, & Wallbott, 2001).

Le traitement cérébral des deux types d'émotions vocales décrites (prosodie émotionnelle et vocalisation émotionnelle) a été comparé expérimentalement à l’aide de l’électro-encéphalographie. Pell et collaborateurs (2015) ont décrit un traitement préférentiel des émotions provenant de vocalisations émotionnelles non verbales comparées à la prosodie émotionnelle. Plus spécifiquement, ils ont observé que l’activité cérébrale associée à ces deux types de stimuli permettait de les différencier. Les vocalisations émotionnelles étaient traitées de manière plus précoce (N100) et provoquaient des réponses (P200) plus tôt et plus forte que la prosodie émotionnelle.

Cette rapidité de traitement combiné à leur caractère universel, supporte l’idée (Belin et al., 2004; Belin, Fillion-Bilodeau, & Gosselin, 2008) que ces interjections vocales émotionnelles sont l'équivalent auditif des expressions faciales émotionnelles de base décrites par Ekman (1992).

1.1.2 Les émotions musicales - l'art de transmettre des émotions

Parallèlement à la voix, nous sommes également constamment exposés à la musique dans notre environnement quotidien. Des études (Juslin, Liljeström, Västfjäll, Barradas, & Silva, 2008; North, Hargreaves, & Hargreaves, 2004) utilisant des techniques d'évaluation écologique momentanée (p.ex. des envois à intervalles aléatoires de questions, par message texte, au cours de la journée) ont permis de documenter que la musique était présente lors de 30-40% de ces envois et que celle-ci pouvait influencer l'humeur du répondant (Juslin et al., 2008). Cette capacité de la musique à induire et réguler les émotions est la raison la plus citée dans la population générale pour expliquer l'écoute de musique (Juslin & Sloboda, 2001; Lonsdale & North, 2011). Généralement, bien que les émotions musicales puissent différer, la plupart des adultes en ressentent lorsqu'ils écoutent de la musique (Zentner, Grandjean, & Scherer, 2008).

Selon de nombreux chercheurs, les émotions musicales de base seraient la joie, la tristesse et la peur, car ces émotions sont parmi les plus faciles à reconnaitre et à communiquer dans la musique (Gabrielsson & Juslin, 2003; Juslin & Laukka, 2003). Une sensibilité à ces émotions musicales de base a même été observée chez les personnes ayant un trouble congénital de perception des hauteurs tonales (Gosselin, Paquette, & Peretz, 2015) et chez les utilisateurs d'implants cochléaires (IC; Ambert-Dahan, Giraud, Sterkers, & Samson, 2015).

Dans la population générale, la perception d'émotions dans la musique est similaire à un réflexe, les adultes ont besoin de moins d'un quart de seconde de musique (un accord ou quelques notes) pour catégoriser les extraits musicaux comme étant joyeux ou tristes (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Filipic, Tillmann, & Bigand, 2010; Peretz, Gagnon, & Bouchard, 1998).

Chez les nouveau-nés (1-3 jours), l'activité cérébrale liée à l'écoute de musique permet de déterminer que leur architecture neurale, entre autres les structures limbiques, est sensible aux changements de musique consonante (agréable) et dissonante (désagréables), supportant l'idée d'une prédisposition biologique aux émotions musicales (Perani et al., 2010). De plus, il a été démontré en utilisant une procédure d'habituation que dès neuf mois les jeunes enfants peuvent facilement discriminer la musique joyeuse et triste (Flom, Gentile, & Pick, 2008). Il a été également possible d'observer que dès trois ans, les enfants démontrent une capacité à reconnaitre la joie dans de la musique complexe et que dès l'âge de six ans, ils montrent des capacités similaires aux adultes pour identifier la tristesse, la peur, la colère dans la musique (Terwogt & Van Grinsven, 1991). Cette capacité reste pratiquement inchangée au cours de la vie (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001).

Similairement à ce qui a été observé pour la voix, si nous avons des prédispositions biologiques à la reconnaissance des émotions musicales, nous devrions être en mesure de déduire les émotions exprimées dans une culture à laquelle nous n'avons jamais été exposés. Il a été démontré que les Mafa, membres d'une tribu africaine isolée, pouvaient reconnaître les émotions dans la musique occidentale au-dessus du hasard (Fritz et al., 2009). Des résultats similaires de reconnaissance d'émotions musicales ont été observés chez des Japonais à qui ont présentait de la musique hindoue et occidentale (Balkwill, Thompson, & Matsunaga, 2004). L'origine de l'interprète a même été contrôlée dans une étude transculturelle (Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013). Des musiciens provenant de cultures musicales différentes (musiques : suédoise folk, hindoustani classique, japonaise traditionnelle, et occidentale classique) ont interprété les mêmes pièces en exprimant 11 émotions. Par la suite, ces extraits ont été jugés par des participants de ces différents groupes culturels. Il a été démontré que toutes les expressions émotionnelles pouvaient être reconnues par tous les groupes au-dessus du niveau du hasard. De meilleures performances lors de l’identification de l'émotion exprimée ont été observées : pour les stimuli provenant des traditions musicales auxquels le groupe de participants appartenait et pour les stimuli exprimant des émotions de base.

En somme, ces différentes études mettent en évidence l'existence d'une certaine invariance dans l'expression des émotions de base à travers les cultures (voir Juslin, 2012, pour une réflexion sur le sujet). L’expertise musicale varie grandement dans nos sociétés, mais il a également été observé que ces jugements émotionnels sont aussi cohérents entre les auditeurs d’une même culture qui varient considérablement en termes de formation musicale (Vieillard et al., 2008).

En résumé, La perception des émotions vocales et musicales se réalise rapidement, se développe très tôt et traverse les cultures. L’humain dans les deux cas semble avoir des prédispositions biologiques qui rendent la perception d’émotions musicales et vocales de base naturelle et sans effort.

1.2 Indices acoustiques des émotions auditives

Depuis que Spencer (1857) a proposé dans son texte ‘‘The origin and function of music” que la musique s’est développée *à partir des rythmes et cadences de discours passionnés*, plusieurs ont souscrit à l’idée que la capacité de la musique à transmettre des émotions est liée à sa capacité à imiter les émotions vocales.

Au niveau empirique, une méta-analyse par Juslin et Laukka (2003) a permis de déterminer que la voix et la musique semblent utiliser les mêmes indices acoustiques pour transmettre des émotions. Leur revue de 145 articles a permis d'identifier dans les deux domaines quels indices acoustiques sont généralement utilisés ou manipulés pour produire une émotion spécifique. Par exemple, dans les deux domaines la joie est généralement associée à un tempo rapide, une intensité moyenne, un niveau moyen de hautes fréquences, une fréquence fondamentale haute (avec une grande variabilité), un contour ascendant et une attaque rapide. En somme, ils ont déterminé que dans les deux domaines plusieurs émotions avaient des profils acoustiques très similaires. De manière analogue, lors d'une étude récente, Curtis et Bharucha (2010) ont démontré que les expressions vocales de tristesse et de colère utilisent des patrons de hauteur tonale similaires à celles utilisées pour transmettre ces mêmes émotions en musique. En règle générale, à travers les différentes études, des grandes similarités acoustiques sont observées et se traduisent généralement par des jugements émotionnels similaires. Certains indices acoustiques influenceraient même directement les jugements émotionnels auditifs (Gosselin et al., 2015 - Annexe 1; Juslin & Laukka, 2003; Quarto, Blasi, Pallesen, Bertolino, & Brattico, 2014). Par exemple, certains indices liés au timbre (la clarté, l'énergie et la rugosité), aux hauteurs tonales et à l'information temporelle (p.ex. le rythme) sont connus pour être directement corrélés avec les jugements émotionnels des participants.

Une étude a manipulé directement certains de ces indices acoustiques pour observer leurs effets sur la perception émotionnelle dans la musique et la voix. Ilie et Thompson (2006) ont trouvé des effets similaires au niveau des réponses émotionnelles à des stimuli musicaux et vocaux en manipulant certains indices acoustiques (intensité, tempo). Les stimuli manipulés pour être plus rapides (musicaux et vocaux) ont été jugés comme ayant plus d'énergie que leurs opposés manipulés pour être plus lents, et les stimuli présentés à plus fort volume ont généralement été perçus comme étant plus énergétiques que leurs opposés présentés à plus faible volume. Renforçant l'idée que certains indices acoustiques ont des effets similaires pour les stimuli vocaux et musicaux. Cependant, ils ont également noté que certaines manipulations (p.ex. hauteurs tonales) affectaient différemment les stimuli vocaux et musicaux; la manipulation de la hauteur tonale (2 demi-tons) des extraits a eu des conséquences opposées pour la musique et la voix. Ce sont les voix plus aigües et les musiques plus graves qui ont été jugées respectivement plus agréables dans leur catégorie. Ces différences pourraient refléter un traitement distinct pour certaines spécificités acoustiques de certaines émotions musicales et vocales (un raffinement émotionnel spécifique à certains timbres) ou encore refléter des différences dans la complexité ou la sorte de stimuli utilisés (discuté plus en détail dans la section 1.4).

Une autre façon d'évaluer l'impact de certains paramètres acoustiques sur la perception en psychologie outre celle des stimuli manipulés est d'avoir recours à des populations spéciales ayant de la difficulté à percevoir certains de ces paramètres, par exemple, les amusiques, des individus qui ont un trouble congénital de perception des fines hauteurs tonales. Leur perception altérée des hauteurs tonales affecte leur reconnaissance de pièces musicales sans l'aide de paroles (Peretz et al., 1998), mais les amusiques performent de manière similaire aux contrôles lorsqu'il est question de reconnaissance d'émotions (Gosselin et al., 2015). Malgré leur déficit, les amusiques peuvent reconnaitre les émotions musicales, mais ils ne sont pas influencés de manière identique aux contrôles dans leurs jugements par les paramètres acoustiques. Comme l'ont pouvait s'en attendre, les amusiques ne semblent pas influencés dans leur jugement par les fines différences de hauteurs tonales, mais ils se sont montrés sensibles aux variations de timbre (aux fluctuations: d'énergie, de clarté et de rugosité). Une autre population chez qui le déficit de perception des hauteurs tonales est encore plus prononcé, et chez qui les impacts de ce déficit sont encore plus grands, sont les utilisateurs d'implants cochléaires (IC).

1.2.1 Conséquences chez les utilisateurs d'implants cochléaires

Les IC peuvent avec succès restaurer l'audition chez des individus profondément sourds. Après des séances intensives de réadaptation, la plupart des utilisateurs atteignent un bon niveau de compréhension du langage. Par contre, le signal acoustique communiqué par l'implant est fortement dégradé, et la résolution des fréquences sonores y est faible. En conséquence, les utilisateurs d'implant cochléaire ont de grandes difficultés à percevoir les patrons de hauteurs tonales (Gfeller & Lansing, 1992; Gfeller & Lansing, 1991; Hopyan, Peretz, Chan, Papsin, & Gordon, 2012) et les changements et/ou la direction d'un changement de hauteurs tonales (Gfeller, Turner, et al., 2002; Laneau, Wouters, & Moonen, 2004) lorsque comparé à l'auditeur normal.

Comprendre les émotions vocales et musicales nécessite le traitement d'indices acoustiques précis (Juslin & Laukka, 2003), dont la plupart sont liés aux hauteurs tonales. En conséquence, les utilisateurs d'IC ont de la difficulté à bien percevoir ces expressions qu'il s'agisse de prosodie ou de musique (Nakata, Trehub, & Kanda, 2012; Wang, Trehub, Volkova, & van Lieshout, 2013). Heureusement, plusieurs indices acoustiques ne reposent pas uniquement sur des variations de hauteurs tonales et peuvent être utilisés pour transmettre des émotions (Gabrielsson & Lindström, 2001; Gosselin et al., 2015) et ces indices acoustiques (p.ex. le rythme et le timbre) peuvent être généralement perçus par les utilisateurs d'IC (Gfeller, Witt, Mehr, & Woodworth, 2002; Kong, Cruz, Jones, & Zeng, 2004; Looi, Gfeller, & Driscoll, 2012).

Cette capacité à utiliser certains indices acoustiques est probablement responsable de la sensibilité aux émotions auditives préservée chez les utilisateurs d'IC. Les utilisateurs d'IC peuvent discriminer/identifier certaines émotions vocales et musicales (joie, tristesse, peur) au-dessus du niveau du hasard, mais pas aussi bien que les auditeurs normaux (Hopyan et coll., 2012; Stabej et coll., 2012; Volkova et coll., 2013, Ambert-Dahan, 2015).

Cette perception dégradée des émotions dans le langage peut nuire à l'intégration sociale et la communication interpersonnelle des utilisateurs d'IC. La perception émotionnelle est d'autant plus cruciale pour le développement des enfants recevant un implant, car le fait de ne pas percevoir les émotions de leurs parents ou de leurs enseignants entraine un comportement inadapté et une incapacité d'exprimer leurs émotions dans leur propre voix. Ce déficit auditif émotionnel peut s'étendre au traitement général des émotions; Wiefferink et collaborateurs (2012) ont observé une perturbation de la régulation émotionnelle (expression de l'émotion et stratégies d'adaptation) et du fonctionnement social (compétence sociale et comportements d'externalisation) chez les utilisateurs d'IC comparativement aux sujets ayant une audition normale. L'accès aux émotions musicales est également important pour les utilisateurs d'IC: de nombreuses personnes sourdes spécifient l'appréciation de la musique comme une motivation majeure pour obtenir un implant (Gfeller, Turner, et al., 2002).

Les études réalisées à ce jour sur la perception émotionnelle musicale et vocale chez les utilisateurs d'IC ne permettent pas d'identifier quels indices acoustiques ils utilisent pour identifier les émotions auditives. Leur perception des émotions musicales et vocales pourrait être grandement améliorée, si les indices acoustiques émotionnels préservés étaient identifiés et utilisés lors de leur réadaptation et/ou mis de l'avant lors de la conceptualisation des logiciels liés aux IC.

Des paramètres acoustiques similaires à ceux utilisés par les amusiques pourraient être possiblement responsables de la sensibilité aux émotions préservées chez les utilisateurs d'IC. Le code combinatoire des paramètres acoustiques responsable de la perception des émotions vocale et musicale reste à être exploré dans cette population, mais somme toute, les résultats de méta-analyse et d'études dans la population générale semblent indiquer une grande similarité acoustique entre les émotions vocales et musicales.

1.3. Corrélats neuronaux des émotions auditives

Les prédispositions biologiques observées pour la perception des émotions vocales et musicales combinées à leur grande similarité acoustique ont incité de nombreux chercheurs en neuroscience à postuler l'existence de substrats neuropsychologiques communs pour traiter les émotions musicales et vocales. Une théorie probable est que la musique recycle les circuits émotionnels qui ont évolué principalement pour le traitement des vocalisations biologiquement importantes (Peretz, Aubé, & Armony, 2013). Les caractéristiques acoustiques communes partagées par les deux domaines pourraient permettre un tel recyclage neuronal, car ils exigeraient la même structure ou le même ensemble de neurones nécessaires à leur traitement (Dehaene & Cohen, 2007).

Récemment, Frühholz et collaborateurs (2016) ont proposé un modèle intégratif pour le traitement des sons émotionnels (entre autres vocaux et musicaux) qui impliquerait en son centre l'amygdale et le cortex auditif. Le modèle attribue entre autres à l'amygdale le rôle de détecteur de pertinence (l'analyse de la valence sociale et émotionnelle des sons courts) et au cortex auditif un rôle équivalent complémentaire : le traitement des caractéristiques sonores (acoustiques) qui se développerait dans le temps. Le traitement émotionnel serait le résultat d'une communication constante entre ces deux régions centrales au modèle et leurs interactions avec les lobes frontaux, les ganglions de la base et le cervelet.

Plusieurs chercheurs ont étudié les régions cérébrales impliquées lors du traitement des émotions vocales ou musicales. L'existence de substrats neuropsychologiques communs devrait se traduire en une grande similarité entre les résultats observés dans les deux cas. J'aborderai d'abord les corrélats neuronaux associés respectivement aux émotions vocales et musicales suivies par la présentation d'études qui ont comparé directement les activations cérébrales liées au traitement des émotions vocales et musicales.

1.3.1 Émotions vocales

Une revue (Schirmer & Kotz, 2006) des études d'imagerie réalisées sur le traitement cérébral de la prosodie émotionnelle met de l'avant que la prosodie engage un réseau distribué de régions : le Sulcus Temporal Supérieur (STS) pour le traitement primaire de l'information auditive, le Gyrus Temporal Supérieur (GTS) et la portion antérieure du STS pour l'intégration des signaux acoustiques liés aux émotions et les gyrus frontaux inférieurs (GFI) et les cortex orbitaux frontaux (COF) pour le jugement émotionnel. Plus récemment, à l'aide de technique d'Imagerie par Résonance Magnétique fonctionnelle (IRMf), il a même été possible de déterminer chez les sujets normaux que les émotions vocales peuvent être différenciées au niveau cortical (Kotz, Kalberlah, & Bahlmann, 2013). À l'aide de l'analyse de patrons multivoxels (MVPA)[2], les auteurs ont déterminé que dans certaines régions cérébrales l'observation de patrons d'activation pouvait permettre de prédire l'émotion présentée au sujet. Les prédictions étaient significativement supérieures au hasard dans certaines régions de l'hémisphère droit: la partie postérieure du GTS, la partie antérieure du STS, le cingulum, l'insula antérieure et l'opercule frontal adjacent, le GFI, et le gyrus frontal médian. L'encodage dans l'hémisphère gauche était au-dessus du hasard dans la partie antérieure et postérieure du gyrus temporal médian.

Au niveau sous-cortical, il a été observé par Fecteau et collaborateurs (2007) que l'amygdale peut-être activée lors de la présentation d'émotions vocales positives (plaisir et rires) et lors d'émotions vocales négatives (pleurs et tristesse). Le rôle de l'amygdale pour le traitement des émotions vocales a aussi été étudié chez des patients cérébrolésés. Il a été observé dans plusieurs cas qu'une lésion de l'amygdale peut compromettre la reconnaissance d'expressions vocales de peur (Dellacherie, Hasboun, Baulac, Belin, & Samson, 2011; Scott et al., 1997; Sprengelmeyer, Rausch, Eysel, & Przuntek, 1998).

Il est à noter que l'implication de l'amygdale n'est pas systématique. Dans certains cas où les patients avaient des lésions bilatérales sélectives à l'amygdale, leur reconnaissance des expressions vocales de la peur était préservée, mais leur reconnaissance des visages de peur était altérée (Adolphs & Tranel, 1999; Anderson & Phelps, 1998). Également certaines études

---

[2] Cette approche utilise une technique de classification des patrons d'activation pour extraire le signal qui est présent dans les activations à travers plusieurs voxels, même si (considérés individuellement) les voxels peuvent ne pas être sensibles de manière significative à l'une des conditions d'intérêt (Norman, Polyn, Detre, & Haxby, 2006). L'analyse MVPA est considérée comme un problème de classification supervisée où un classificateur tente de saisir les relations entre le patron spatial de l'activité IRMf et les conditions expérimentales (Mahmoudi et al., 2012). Le patron multi-voxels de la réponse peut être considéré comme un code combinatoire avec une très grande capacité pour représenter les distinctions entre les états cognitifs.

d’imagerie ne montrent pas d’activation de l'amygdale lors de présentations d’expressions de peur (Morris, Scott, & Dolan, 1999; Pourtois, de Gelder, Bol, & Crommelinck, 2005; Royet et al., 2000). Le fait que l’amygdale s’active pour les émotions positives et négatives, mais pas de manière systématique porte à penser que l'amygdale pourrait agir comme un «détecteur de pertinence» pour les évènements biologiquement significatifs, indépendamment de leur valence (Sander, Grafman, & Zalla, 2003).

1.3.2 Émotions musicales

Différentes structures corticales sont connues pour être impliquées dans le traitement émotionnel de la musique. Les COF (Blood & Zatorre, 2001; Khalfa, Schon, Anton, & Liégeois-Chauvel, 2005; Menon & Levitin, 2005), le cortex temporal supérieur, et le cortex cingulaire antérieur (Blood & Zatorre, 2001; Green et al., 2008; Mitterschiffthaler, Fu, Dalton, Andrew, & Williams, 2007) ont fréquemment été rapportés comme étant liés au traitement des émotions musicales. Récemment, Aubé et collaborateurs (2015), ont également observé des activations spécifiques aux musiques de peur et de joie dans le GTS lorsque celles-ci sont individuellement contrastées à des extraits neutres.

Pour ce qui est des structures sous-corticales, elles semblent également impliquées dans les réponses émotionnelles de la musique. Grâce à une combinaison de techniques d’imagerie, Salimpoor et collaborateurs (2011), ont pu observer chez leurs participants, lors du point culminant du plaisir émotionnel musical, la libération de dopamine endogène dans le striatum et une grande activation du noyau accumbens. Le noyau accumbens est également impliqué dans la réponse à des stimuli très excitants ou motivants (Cooper & Knutson, 2008; Knutson & Cooper, 2005), comme le chocolat (Small, Zatorre, Dagher, Evans, & Jones-

Gotman, 2001) et l’abus de drogue/cocaïne (Breiter et al., 1997). Ainsi, la musique peut être aussi efficace que la nourriture, la drogue et les expressions du visage pour moduler les réponses émotionnelles sous-corticales.

Également, comme pour la voix, la capacité de percevoir des émotions musicales spécifiques peut être perdue après une lésion cérébrale. C'est le cas pour la reconnaissance de la peur et la tristesse après des atteintes à l'amygdale (Gosselin, Peretz, Noulhiane, & Hasboun, 2005; Gosselin, Peretz, Johnsen, & Adolphs, 2007).

1.3.3 Comparaison des émotions vocales et musicales

Deux études ont directement comparé les corrélats neuronaux associés aux émotions musicales et vocales et ont observé des régions cérébrales associées spécifiquement au traitement de la musique et de la voix et peu de résultats spécifiques au traitement émotionnel. Escoffier et collaborateurs (2013) ont utilisé l’IRMf pour comparer les activations liées aux émotions vocales et musicales (joie, tristesse et neutralité) lors de tâches explicite et implicite de reconnaissance d’émotions. Lors de leurs analyses à travers tout le cerveau, ils ont observé des activations spécifiques aux stimuli vocaux dans les cortex temporaux supérieurs et moyens et aux stimuli musicaux dans le planum temporale et le gyrus d’Heschl, bilatéralement. Le contraste des tâches explicite et implicite de reconnaissance d’émotions a été associé à une activation dans le lobe frontal. Aucun effet n’a été observé pour les contrastes des émotions versus les stimuli neutres, ils ont dû se résoudre à l’utilisation de régions d’intérêt (ROI) pour explorer ses contrastes émotionnels et ont découvert une activation plus grande pour les stimuli exprimant la joie que ceux exprimant la tristesse et la neutralité dans le planum temporale et le gyrus d’Heschl. Leurs résultats suggèrent un traitement spécifique pour certaines caractéristiques des stimuli vocaux et musicaux, mais sous-entendent également des mécanismes similaires pour le traitement des émotions vocales et musicales, particulièrement pour la joie.

Aubé et collaborateurs (2015) ont également comparé les émotions vocales et musicales, et avaient inclus une comparaison avec des expressions faciales. À traves tous les stimuli (visuels et auditifs), ils ont observé un effet principal lié aux stimuli de peur (vs neutralité) au niveau de l'amygdale/hippocampus et de la partie postérieure de l'insula. Ils n'ont pas présenté de données contrastant seulement la musique, la voix et leurs émotions (joie, tristesse, peur, neutralité), mais ils ont contrasté globalement les stimuli vocaux et musicaux et ont identifié une activité spécifique aux voix dans le STS et aux musiques notamment dans le GTS. Ils ont également contrasté les stimuli présentés de manière auditive aux visages et ont observé une différence dans le GTS. En somme, leurs résultats évoquent également un traitement spécifique pour certaines caractéristiques des stimuli vocaux et musicaux, mais suggèrent aussi des mécanismes similaires pour le traitement des émotions visuelles, vocales et musicales, particulièrement pour la peur.

Bien que la convergence à travers de méthodes différentes soit souhaitable et que ces études (Aubé et al., 2015; Escoffier et al., 2013) nous ont beaucoup appris sur le sujet, leur choix de stimuli pourrait être au centre des différences observées entre les stimuli vocaux et musicaux. La première étude (Escoffier et al., 2013) a utilisé de longs (> 10 secondes) stimuli: des extraits complexes de musique familière et des stimuli vocaux composés d'énonciations multiples de voyelles décousues enchainées avec une prosodie émotionnelle. Dans la deuxième étude (Aubé et al., 2015), de courts extraits de musique écrits dans le système tonal occidental (certains sur un piano; sans progression continue entre les notes comme dans la voix) ont été contrasté à des vocalisations non contrôlées pour leur structure segmentée (exprimées sur différents sons).

Dans ses deux études d'imagerie (Aubé et al., 2015; Escoffier et al., 2013), les différences d'activation cérébrales pour les stimuli vocaux et musicaux pourraient refléter un traitement différent pour les voix et musiques émotionnelles, ou pourraient refléter des différences entre les stimuli utilisés.

1.4. Limites des comparaisons des émotions auditives

Plusieurs types de stimuli musicaux et vocaux ont été utilisés pour étudier les émotions évoquées de manière auditive. Des exemples de stimuli musicaux utilisés incluent des pièces musicales longues issues de musiques populaires (p.ex. Escoffier et al., 2013) ou écrites selon les règles du système tonal occidental (p.ex. Dalla Bella et al., 2001; Vieillard et al., 2008) et des extraits de musique plus courts (p.ex. Aubé et al., 2015). Des exemples de stimuli vocaux utilisés incluent des éclats émotionnels non verbaux (p.ex. Belin et al., 2008; Pell et al., 2015) et des mots ou des phrases prononcés avec une prosodie émotionnelle (p.ex. Banse & Scherer, 1996; Buchanan et al., 2000; Kotz, Meyer, Alter, Besson, & Cramon, 2003; Pell, 2006; Schirmer, Striano, & Friederici, 2005; Schirmer & Kotz, 2006).

Bien que ces stimuli aient été très utiles pour étudier les émotions, trop de caractéristiques différencient ces stimuli musicaux et vocaux, les rendant ainsi difficiles à comparer dans une étude contrôlée. Cela est particulièrement vrai pour les facteurs tels que la durée, le niveau de complexité ainsi que le contexte dans lequel les stimuli ont été créés. Ces différences nuisent à une comparaison directe que ce soit pour les études behaviorales chez les auditeurs normaux ou les études chez les populations spéciales.

Tout en maximisant la valeur écologique de l'étude, l'utilisation de vrais morceaux de musique peut introduire une variabilité non contrôlée de nombreux paramètres : acoustiques (les relations temporelles ou tonales établies dans la culture occidentale), attentionnelles, et mnésiques. Ces différences acoustiques et cognitives sont susceptibles de recruter différents réseaux de neurones (Peretz & Zatorre, 2005), ce qui rend l'interprétation des zones corticales activées difficiles. Parmi toutes les zones corticales mentionnées précédemment, il n'est pas toujours facile de déterminer si l'activité est liée au traitement émotionnel ou non émotionnel de la structure musicale ou si celle-ci est spécifique aux stimuli vocaux utilisés. D'où l'importance d'une bonne sélection des stimuli pour les études d'imagerie comparant les émotions vocales et musicales.

Des travaux ont été réalisés pour valider des stimuli musicaux et vocaux, mais jamais avec l'objectif d'être directement comparé. Le domaine des émotions auditives à un grand besoin de stimuli vocaux et musicaux qui sont épurés et comparables, de manière à pouvoir étudier les différences liées au traitement de ses deux types de stimuli auditifs.

Récemment, un ensemble validé d'éclats émotionnels vocaux, conçus comme une contrepartie auditive des visages d'Ekman a été validé (Belin et al., 2008). Les Voix Affectives Montréalaises (MAV) sont constituées d'un ensemble de courtes interjections vocales (sur la voyelle /ɑ/) exprimant la colère, le dégout, la peur, la douleur, la tristesse, la surprise, la joie, le plaisir sexuel et la neutralité. Les interjections ont été enregistrées par des acteurs professionnels (5 femmes, 5 hommes), chaque acteur a exprimé chaque emotion, pour éviter un effet de l'identité de l'acteur. Les MAV ne contiennent aucune information linguistique-sémantique. Elles représentent de courtes expressions primitives de ces émotions. Les MAV sont des stimuli qui permettent l'étude des mécanismes psychologiques qui sous-tendent le traitement auditif affectif avec un minimum d'interaction avec les processus linguistiques (Bestelmeyer, Rouger, DeBruine, & Belin, 2010).

Ces stimuli seraient donc très appropriés pour être comparés à des stimuli émotionnels musicaux, si leurs équivalents musicaux étaient créés et validés.

1.5. Objectifs

L'objectif de cette thèse est d'apporter un support à l'existence de substrats neuropsychologiques communs pour le traitement des émotions musicales et vocales; à l'hypothèse que la musique recycle les circuits émotionnels qui ont évolué principalement pour le traitement des vocalisations biologiquement importantes (Isabelle Peretz, 2010). Le support empirique à cette théorie est apporté par la comparaison en profondeur des émotions musicales et vocales en minimisant les facteurs qui pourraient différencier les deux types de stimuli. Cette comparaison est présentée à travers 3 articles scientifiques.

1.5.1 Validation de stimuli

Article 1: Ce premier article publié dans *Frontiers in emotion science* (Paquette, Peretz, & Belin, 2013), avait comme objectif de répondre au besoin de stimuli comparables exprimant des émotions vocales et musicales. Une batterie de stimuli musicaux conçus pour être l'équivalent des Voix Affectives Montréalaises y est introduite. Des éclats émotionnels musicaux (MEB) exprimant la peur, la joie la tristesse et la neutralité ont d'abord été enregistrés et un processus de validation a été réalisé pour déterminer si ces brèves expressions musicales émotionnelles pouvaient être correctement identifiées. Par la suite, celles-ci ont été comparées sur différents jugements émotionnels (valence, arousal) aux stimuli vocaux, pour déterminer s'ils étaient similaires.

1.5.2 Comparaison des corrélats neuronaux des émotions auditives

Article 2: Ce deuxième article avait comme objectif d'apporter un support au postulat selon lequel la musique et la voix auraient des circuits émotionnels communs. En minimisant les facteurs qui pourraient différencier les deux types de stimuli, nous voulions minimiser l'activité cérébrale spécifique aux stimuli et identifier les régions limbiques et corticales communes à la voix et la musique qui sont spécifiques au traitement émotionnel. Une étude en Imagerie par Résonnance Magnétique fonctionnelle (IRMf) a été réalisée à l'aide des stimuli vocaux et musicaux émotionnels développés et comparés dans l'article 1.

1.5.3 Comparaison des émotions auditives chez les utilisateurs d'IC

Article 3: Ce troisième article, à volet plus clinique, avait comme objectif d'identifier, les indices acoustiques vocaux et musicaux qui sont utilisés par les utilisateurs d'IC pour réaliser des jugements émotionnels. Certains indices acoustiques (p.ex. le rythme et le timbre) peuvent être perçus par les utilisateurs d'IC. Si les indices acoustiques associés au jugement émotionnel chez les utilisateurs d'IC peuvent être identifiés, ils pourraient faire l'objet d'une plus grande attention lors du développement des logiciels d'implants ou lors de thérapie de réadaptation. Pour se faire, les stimuli vocaux et musicaux très similaires, développés dans l'article 1, ont été présentés à des auditeurs normaux, avec et sans simulation[3] d'IC.

[3] La simulation d'implant cochléaire est une approche validée qui permet de tester l'écoute normale et avec implant chez les mêmes participants. Les algorithmes de codage vocal ont été utilisés avec succès pour simuler la dégradation du signal auditif par un IC chez les sujets ayant une audition normale; leur utilisation pour la recherche a été validée pour la parole (Fu & Nogaki, 2005; Nogaki et al., 2007; Poissant et al., 2006; Qin & Oxenham, 2003) et la musique (Cousineau et al., 2010).

Chapitre 2: Méthodologie et Résultats

A1: A validated set of musical affect bursts to investigate auditory affective processing

Publié dans Frontiers in Psychology: Emotion Science (2013), volume 4 (509), pages 1-7.

Sébastien Paquette[1], Isabelle peretz[1], Pascal Belin[1,2,3]

[1] Department of Psychology, International Laboratory for Brain Music and Sound Research, Center for Research on Brain, Language and Music, University of Montreal, Montreal, QC, Canada

[2] Voice Neurocognition Laboratory, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

[3] Institut des Neurosciences de La Timone, Aix-Marseille Université, Marseille, France

Keywords: Music, Emotion, Auditory stimuli, Voice

The Musical Emotional Bursts

Abstract

The Musical Emotional Bursts (MEB) consist of 80 brief musical executions expressing basic emotional states (happiness, sadness and fear) and neutrality. These musical bursts were designed to be the musical analog of the Montreal Affective Voices (MAV) - a set of brief non-verbal affective vocalizations portraying different basic emotions. The MEB consist of short (mean duration: 1.6s) improvisations on a given emotion or of imitations of a given MAV stimulus, played on a violin (10 stimuli × 4 [3 emotions + neutral]), or a clarinet (10 stimuli × 4 [3 emotions + neutral]). The MEB arguably represent a primitive form of music emotional expression, just like the MAV represent a primitive form of vocal, non-linguistic emotional expression. To create the MEB, stimuli were recorded from 10 violinists and 10 clarinetists, and then evaluated by 60 participants. Participants evaluated 240 stimuli [30 stimuli × 4 (3 emotions + neutral) × 2 instruments] by performing either a forced-choice emotion categorization task, a valence rating task or an arousal rating task (20 subjects per task); 40 MAVs were also used in the same session with similar task instructions. Recognition accuracy of emotional categories expressed by the MEB (n:80) was lower than for the MAVs but still very high with an average percent correct recognition score of 80.4%. Highest recognition accuracies were obtained for happy clarinet (92.0%) and fearful or sad violin (88.0% each) MEB stimuli. The MEB can be used to compare the cerebral processing of emotional expressions in music and vocal communication, or used for testing affective perception in patients with communication problems.

Introduction

With increasing knowledge in the field and new methods to explore the human brain, emotions are no longer too obscure or subjective to be studied scientifically. In neuroscience, many research projects are now entirely dedicated to the study of emotion. Thus, it appears timely to construct a standardized and validated set of stimuli and to make these freely and easily available (http://vnl.psy.gla.ac.uk) in order to facilitate the comparability of future studies.

A great amount of work has been achieved in the field of visually perceived emotions, utilizing validated stimuli like the International Affective Picture System and the Ekman faces (Dailey, Cottrell, & Reilly, 2001; Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002; Lang, Öhman, & Vaitl, 1988), which were designed to portray basic emotions (anger, disgust, fear, happiness, sadness, and surprise as well as a neutral expression). These validated sets of stimuli have provided highly useful tools for the study of brain structures (e.g. amygdala: Adolphs, Tranel, Damasio, & Damasio, 1994) involved in emotional processing and its developmental trajectory (Charlesworth & Kreutzer, 1973). With the same objectives, an increasing number of studies are being conducted in the domain of aurally perceived emotions, thus calling for validated stimuli sets.

A large part of the research on auditory affective processing has been conducted on speech prosody utilizing words or sentences spoken with various emotional expressions (Banse & Scherer, 1996; Buchanan et al., 2000; Kotz et al., 2003; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Monrad-Krohn, 1963; Pell, 2006; Schirmer et al., 2005). Another way to express an emotion vocally is via non-verbal affect bursts (Scherer, 1994; also sometimes called non-verbal interjections: Schröder, 2003). Non-verbal affect bursts are vocal expressions (e.g., screams, laughter) that usually accompany intense emotional feelings. Affect bursts are minimally conventionalized, thus a relatively universal means of spontaneous human communication (see Koeda et al., 2013; Sauter et al., 2010, for cross-cultural studies). They are believed to reflect more of a biological push than a sociological pull (Scherer, 1986); they are closer to the primitive affect expressions of babies and animals than to emotional speech.

Recently, a validated set of auditory affect bursts designed as an auditory counterpart of Ekman faces was recorded and validated by Belin et al. (2008). The so-called Montreal Affective Voices (MAV) consist of a set of short vocal interjections on the vowel /a/ expressing anger, disgust, fear, pain, sadness, surprise, happiness, sensual pleasure, and neutrality. The MAV represent short primitive expressions of these emotions with minimal semantic information, providing useful stimuli for the study of the psychological mechanisms underlying auditory affective processing with minimal interaction with linguistic processes (e.g. Bestelmeyer et al., 2010).

However, vocal affect bursts are not the only means of transmitting auditory emotions. Music is often described as the "language of emotions," and recent research on basic musical emotions has shown that emotion recognition in music is consistent across listeners (Vieillard et al., 2008). The terms "basic emotions" correspond to a limited number of innate and universal emotion categories (happiness, sadness, anger, fear, and disgust) from which all other emotions can be derived (Ekman, 1982). Moreover, many studies have demonstrated that emotions in music fit Ekman's definition of basic emotions, they are recognized quickly [only a quarter of a second of music; one chord or a few notes (Bigand, Vieillard, et al., 2005;

Peretz et al., 1998)], early in development (Flom et al., 2008; Terwogt & Van Grinsven, 1991), and across different cultures (Balkwill, Thompson, & Matsunaga, 2004). The latter is even true for cultures without previous exposure to western music (Fritz et al., 2009).

Perception of specific musical emotions (e.g., fear and sadness) can also be lost after damage to the amygdala (Gosselin, Peretz, Noulhiane, & Hasboun, 2005; Gosselin, Peretz, Johnsen, & Adolphs, 2007), suggesting that damage to the limbic system affects perception of basic musical emotion just as reported for other domains (e.g. facial expression: Adolphs et al., 1994; vocal expression: Dellacherie et al., 2011).

An important question that ensues is why music moves us? Recent studies have shown that certain brain areas [e.g., the striatum (Salimpoor et al., 2011), the amygdala (Gosselin et al., 2007)] are associated with musical emotional processing. These same areas have also been associated with basic biological functions (sex, pain). How can we conceptualize the relationship between music and these neurobiological substrates? One possibility is that music co-opts or invades emotional circuits that have evolved primarily for the processing of biologically important vocalizations [e.g., laughs, screams (Peretz, 2010)]. There is currently little experimental data supporting or invalidating the existence of a common musical and vocal channel.

For example, Lima and Castro (2011), demonstrated that musical expertise enhances the recognition of emotions in speech prosody, suggesting that expertise in one domain could translate to the other. Conversely, Thompson and collaborators (Thompson, Marin, & Stewart, 2012), reported that amusics (individuals with a pitch perception deficit; Peretz, Ayotte, Zatorre, Mehler, & Ahad, 2002) were also impaired in perceiving emotional prosody in speech.

More specifically, Ilie and Thompson (2006) compared domains by evaluating the effect of manipulating acoustic cues common to both the voice and music [intensity, rate (tempo), and pitch height] on emotional judgments. They found that loud excerpts were judged as more pleasant, energetic and tense compared to soft excerpts, and that fast music and speech were judged as having greater energy than slow music and speech. However, it was also found that tempo and pitch had opposite effects on other emotional scales. Their results support the view that the processing of musical and vocal emotion could utilize common circuitry, but that some of this circuitry might be domain specific.

The existence of domain-specific processes for decoding emotion is consistent with neuropsychological dissociations found between music and language (Lima, Garrett, & Castro, 2013; Omar, Henley, Bartlett, & Hailstone, 2011; Peretz & Coltheart, 2003). These dissociations could be explained by the fact that musical emotion needs to be actively decoded by the brain based on associations learned via exposure to a musical culture (Juslin & Västfjäll, 2008; Peretz et al., 1998) and past experience with music (Eschrich et al., 2008); since not all musical emotional acoustic parameters are present in emotional vocalizations (e.g. harmony: Juslin & Laukka, 2003), it is possible that these additional cues require additional processing.

Musical and vocal stimuli have both been used to study auditory perceived emotions (Music: Aubé, Peretz, & Armony, 2013; Roy, Mailhot, Gosselin, Paquette, & Peretz, 2009; Vieillard et al., 2008; Voices: Belin et al., 2008; Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Fecteau, Belin, Joanette, & Armony, 2007; Pell, 2006; Schirmer et al., 2005). Although such stimuli have been quite useful to help exploring aurally perceived emotions in their respective channel, many characteristics set current musical and vocal stimuli apart making them hard to compare in a controlled study. This is especially true for factors such as musical structure (limited by mode or tempo), length, level of complexity as well as the context in which they have been created. The use of pre-existing music can introduce uncontrolled variability of many acoustic parameters, with various demands on attention and memory. Such acoustic and cognitive differences are likely to recruit different neural networks (Peretz & Zatorre, 2005). This is why it is important to create and validate musical stimuli that would be as similar as possible to the MAV to allow for a more proper comparison of aurally (musical and vocal) perceived emotions.

The purpose of the present study is to make available for future research a validated set of brief musical clips expressing basic emotions, designed as a musical counterpart of the MAV. We chose to only include happiness, sadness, and fear because these emotions are among the easiest to recognize from music (Gabrielsson & Juslin, 2003; Juslin & Laukka, 2003; see Zentner, Grandjean, & Scherer, 2008 for a more nuanced range of musically induced emotions).

Brief "musical emotional bursts" (MEB) depicting neutral and emotional (happy, sad, and fear) expressions have been recorded from different musicians. The violin and the clarinet were chosen as instruments, not only because they are representative of two different classes of instruments (strings and woodwind) but also because they share important similarities with the voice: "The quasi-vocal quality implied by a seamless progression between notes is a characteristic that can be cultivated in both the clarinet and the violin" (Cottrell &

Mantzourani, 2006:33). These recordings were then pre-selected and validated based on listeners' emotion categorization accuracy, as well as on valence and arousal ratings.

Materials and methods

Recording

Participants

Twenty professional musicians (10 violinists, 10 clarinettists) participated in the recording sessions, after providing written informed consent. They received a compensation of 20$ per h.

Procedure

The musicians were first instructed to perform 10 short improvisations with different levels of expressiveness. They were not told in advance what the recording session was about; on the day of the recording they were told one after the other the emotion they were supposed to improvise on, [fear (as if they were scared), happiness, sadness, and neutrality]. They were told their improvisation had to last around a second (they could practice with the metronome), when ready they realized 10 renditions of the emotion. Neutral stimuli were presented just like any other category of stimuli, but characterized as “without emotion.” After improvising, the same musicians were asked to imitate one after another four MAV stimuli depicting fear, happiness, sadness, and neutrality; they could listen to the stimuli as often as they wished. If the emotional category of the musical burst was not clearly recognized by the experimenter (SP) or if the improvisations were too long they were discarded.

The musical bursts were recorded in a sound-treated studio using a TLM 103 large diaphragm microphone Neumann (Georg Neumann, Berlin, Germany) at a distance of approximately 30 cm. Recordings were pre-amplified using a Millennia Media HV-3D preamplifier and digitized at a 44-kHz sampling rate at 24-bit resolution, using Apogee AD16X. Subsequently they were edited into short segments and normalized at peak value (90% of maximum amplitude), using Adobe Audition 3.0 (Adobe Systems, Inc. San Jose, CA).

We ended up with more stimuli than expected, because each musician gave us more excerpts than we asked for. In total, 1505 improvisations [a minimum of 10 × 4 emotions (happy, sad, fear, and neutral) per musician] and 319 imitations of the MAV [a minimum of 4 × 4 emotions (happy, sad, fear, and neutral) per musician] were recorded.

Stimulus pre-selection

Improvisations lasting longer than 4 s were excluded. Improvisations or imitations containing an artifact (breathing, vocal sounds, breaking bow hair sounds) were also excluded. In the end, the clearest and most representative stimuli (120 Violin-MEB and 120 Clarinet-MEB) were selected for the validation phase, regardless of their type (improvisation or imitation).

Validation

Participants

Sixty participants (19 males) aged from 19 to 59 years (M: 28.8; SD: 9.2), with normal hearing participated in an on-line validation test. Each participant gave informed consent and filled out a socio-demographic information questionnaire prior to the judgment phase. Fifteen participants had 6 years or more of musical education and 45 had 5 years or less of training. They were compensated 3£ for their participation.

Procedure

Participants were instructed to evaluate each of the 240 MEB and 40 MAV (The MAV were included for comparison with the vocal stimuli). There were 30 violin-MEB, 30 clarinet-MEB and 10 MAV per emotion, and all were presented in a random order. Twenty of the 60 participants performed a four alternative forced-choice identification task “Please choose the emotion you think this stimulus represents” among fear, happiness, sadness, and neutrality labels, 20 participants gave arousal ratings “Please rate on the scale below the perceived arousal of the emotion expressed (from 1 not at all aroused to 9 extremely aroused)” and another group of 20 participants gave valence ratings “Please rate on the scale below the perceived valence of the emotion expressed (from 1 extremely negative to 9 extremely positive).”

Results

The stimuli (40 violin-MEB and 40 clarinet-MEB) that were best identified (by being categorized in the intended emotion) by the largest amount of participants were selected (10 MEB; 7 improvisations, 3 imitations- per emotion). In the presence of identical ratings, the briefest stimuli were selected. Due to the small number of stimuli in each category, improvisations and imitations were not analysed separately (separate Tables can be found in the supplementary Material).

Acoustical analyses were also performed to allow users to individually select their stimuli (see supplementary material).

Emotional categorization

Overall accuracy in the four-alternative emotions categorization task is 85.5% (SD: 15.8) for the violin-MEB, 75.4% (23.9) for the clarinet-MEB, and 94.8% (12.1) for the voice-MAV. The average percentage of correct recognition of each intended emotion for the selected stimuli are presented in Table 1. As can be seen, timbre had a greater effect on certain emotional intentions than on others. For example, fear was more difficult to recognize when expressed on a clarinet than on any other timbre.

[Insert table 1 here]

The ANOVA conducted on the recognition scores (see values in bold in Table 1.) with Timbre (violin, clarinet, and voice) and Emotion (happiness, sadness, fear, and neutrality) as within-subject factors yielded a main effect of timbre [$F(2, 38) = 79.51$, $p < 0.001$, η2 = 0.81] and of emotion [$F(3, 57) = 6.81$, $p < 0.005$, $\eta^2 = 0.26$]; however, they are modulated by a significant interaction between Timbre and Emotion [$F(3.4, 64.4) = 16.41$, $p < 0.001$, $\eta^2 = 0.46$, corrected Greenhouse-Geisser].

Recognition scores were compared using Tukey's honestly significant difference. Scores averaged across emotions for each timbre were all significantly different (all $p < 0.005$) from one another: voices yielded the highest recognition scores and clarinet the lowest. Comparing emotions, fear was overall significantly ($p < 0.01$) less accurately recognized than all other emotions.

Using binomial tests to determine if the emotions conveyed by each of the 80 stimuli were recognized above chance level (25%), we found that 87.5% (70/80) of the MEB were recognized above chance ($p < 0.05$; bonferroni corrected). Thus, most MEB are effective in expressing an emotion on a musical instrument. Eight of the 10 stimuli that failed to be recognized belonged to the clarinet-fear category; the other two stimuli were from the violin-joy category.

Emotional ratings

The arousal and valence ratings averaged across participants for each stimulus are presented in Figure 1. The individual ratings are provided in the Supplementary Material.

[Insert figure 1 here]

The same ANOVA with Timbre and Emotion as between-subjects factors as the one performed on the recognition scores was computed on the arousal ratings. A main effect of Timbre [$F(2, 38) = 10.05$, $p < 0.001$, $\eta^2 = 0.35$] and of Emotion [$F(3, 57) = 33.94$, $p < 0.001$, $\eta^2 = 0.64$] were observed; however as previously an interaction between Timbre and Emotion was obtained [$F(6, 114) = 5.85$, $p < 0.001$, $\eta^2 = 0.24$].

In general, the clarinet stimuli were judged to be less arousing than the violin and the vocal ones (all $p < 0.05$; by Tuckey's tests), whereas the latter two were judged to be equally arousing ($p = 0.67$). Neutral expressions were overall significantly less arousing ($p < 0.001$) than all other emotions, and happy stimuli were found to be more arousing ($p < 0.001$) than the sad ones.

It is important to note that the stimuli played on a clarinet were rated differently than the violin and vocal stimuli. Happy clarinet stimuli were rated as more arousing than all the other emotions played on clarinet (all $p < 0.05$), fear stimuli were also significantly ($p < 0.005$) more arousing then the neutral stimuli. In contrast however, the only significant difference for violin and vocal emotional bursts was that neutral stimuli were significantly less arousing (all $p < 0.01$) than all other stimuli.

Regarding valence ratings, we found qualitatively a similar pattern for both the violin and vocal stimuli (Happy > Neutral > Fear > Sad). The clarinet stimuli showed however a slightly different pattern, where fear was rated as being more positive than neutral stimuli (Happy > Fear > Neutral > Sad). Again, both a main effect of Timbre [$F(2, 38) = 6.13$, $p < 0.05$, $\eta^2 = 0.24$] and of Emotion [$F(3, 57) = 116.65$, $p < 0.001$, $\eta^2 = 0.86$] were observed, while the interaction between Emotion and Timbre was again found to be significant [$F(6, 114) = 31.64$, $p < 0.001$, $\eta^2 = 0.63$]. Overall, violin MEB were judged to be less positive than the vocal ones ($p < 0.005$), but globally emotions were significantly different from one another in terms of their valence ratings ($p < 0.005$).

This interaction can be explained by the fact that some differences were observed within timbre. Among the vocal stimuli, the happy ones were judged to be more positive than the neutral ones which were rated as more positive than fear, which in turn was also rated more positively than sadness (all $p < 0.01$). When played on a musical instrument, the happy stimuli were also judged as most pleasant (all $p < 0.001$), whereas only the sad stimuli were rated as significantly more negative than the neutral ones when played on violin ($p < 0.05$), and also as more negative than the stimuli expressing fear played on the clarinet ($p < 0.005$).

Discussion

Here, we validate the MEB — a set of short music clips designed to express basic emotions (happy, sad, fear, and neutral). Despite their short duration (1.6 s on average), the MEB stimuli were correctly categorized by emotion with high accuracy (average recognition score of 80.4%). The highest accuracy was obtained on the violin for stimuli expressing fear and sadness (88%) and on the clarinet for those conveying happiness (92%). Although, the MAV stimuli were best recognized, the newly created MEB were still accurately portraying the desired emotions.

Only three emotions were tested here to allow for direct comparison between basic vocal (MAV) and musical (MEB) emotions. Our limited selection of emotions does limit voice-music comparison, but it is a first step in making that comparison. We acknowledge that there are multiple declinations of positive and negative emotions in the musical and vocal literature, our aim was to use the most easily recognized common to both domains. From a dimensional approach, basic emotions can be distinguished on the dimensions of valence and arousal; variations of these (and other) emotions also differ in valence and arousal and can easily be represented along basic emotions.

The arousal and valence ratings obtained here fit well with this dimensional representation of emotions, with happy stimuli as conveying positive and arousing emotions, fear stimuli as conveying negative and arousing emotions (with the exception of a few clips played on clarinet), sad stimuli as conveying moderately arousing and negative emotions, and the neutral stimuli as conveying an emotional valence that is neither positive or negative with little arousal.

Although the valence scale had a highest rating possible of 9, it is important to note that the maximal average arousal elicited by our stimuli is 6.8 (7.1 for voice), Perhaps the short duration of our stimuli limited their arousing capabilities and could potentially explain the partial overlap in arousal observed in Figure 1 between our two negative emotions (fear, sadness). Also, the fact that the valence scale ranged from "extremely negative" to "extremely positive" (Aubé et al., 2013; Belin et al., 2008), and not from "unpleasant" to "pleasant" could explain why the sad stimuli are differently positioned on the scale than in previous studies (e.g. Vieillard et al., 2008). Nevertheless, our results are still quite similar to those of Vieillard and collaborators (2008) which were obtained with longer and more conventional musical stimuli (inspired from film music), suggesting that the MEB may tap into similar emotional processes as those evoked by more elaborate film music clips. Yet, the MEB consist of brief expressions and are less likely to involve high-level cognitive mechanisms such as divided-attention and sophisticated knowledge of musical structure than more conventional musical stimuli. The MEB are not limited by tonality or defined by a specific rhythm; they were created as short musical bursts, by professional musicians on their instrument.

Our stimuli can be viewed as a primitive form of musical emotion, situated somewhere in between long musical excerpts from recordings (e.g. Peretz et al., 1998) or short musical segments extracted from these (Dalla Bella, Peretz, & Aronoff, 2003; Filipic et al., 2010) and synthesized frequency-modulated tones designed to mimic key acoustic features of human vocal expressions (Kantrowitz, Leitman, Lehrfeld, & Laukka, 2011). Our novel stimuli were created to be exactly where they are in this spectrum by representing the most basic form of musical emotion that can be closely related to vocal expressions. Although exact replicas of the MAV could have been used instead, by digitally transposing the MAV to another timbre, we chose to produce new recordings in order to keep the stimuli as natural (realistic) as possible.

The timbre, or instrument on which music is played, is known to have an important impact on emotion recognition (Balkwill & Thompson, 1999; Behrens & Green, 1993; A. Gabrielsson & Juslin, 1996; Hailstone et al., 2009) For example, Hailstone et al. (2009) have found that melodies sound less happy when played on the violin than on other instruments, as we found here. This effect was particularly clear in the imitations of vocal expressions (see supplementary material). A range of timbres were used in prior studies (including violin and voice) and each instrument seemed to present its own possibilities and limitations when it came to expressing specific emotions. For instance in our study, we observed that fear was not well recognized when expressed on the clarinet.

Other limitations will also need to be addressed. For example, a forced-choice emotion recognition task was used here, and such tasks can have an impact on statistical analyses, such as increased co-linearity (if one response is chosen, the others are not), which generates artificially high recognition rates (Dunlap & Cornwell, 1994; Frank & Stennett, 2001). This method was selected to facilitate the web-based validation procedure of a large number of stimuli (280), and we believe the technique has served its purpose, as significant differences were observed between the timbres and within each timbre as revealed within the confusion matrix.

In addition, musicians were explicitly asked to imitate vocalizations (3/10 MEB per emotion). Such imitations produced on an instrument with voice-like characteristics may limit the chance to obtain domain-specific responses. In contrast, by using such a setup, finding evidence for domain-specificity would be compelling.

Here we propose a validated set of auditory stimuli designed as a musical counterpart of the MAV to allow a better comparison between auditory (musical and vocal) stimuli designed to convey emotions. We hope that the MEB will contribute to the understanding of emotions across domains and modalities.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This work was supported by an Auditory Cognitive Neurosciences, Erasmus Mundus Mobility Scholarship by the European Union to Sébastien Paquette, by grants from Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research and from the Canada Research Chairs program to Isabelle Peretz and BBSRC grants BB/E003958/1 and BB/I022287/1, and ESRC/MRC large grant RES-060-25-0010 to Pascal Belin. We thank Patrice Voss for editing a previous version of the manuscript.

References

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672. http://doi.org/10.1038/372669a0

Aubé, W., Peretz, I., & Armony, J. L. (2013). The effects of emotion on memory for music and vocalisations. *Memory*, *21*(8), 1–11. http://doi.org/10.1080/09658211.2013.770871

Balkwill, L.-L., & Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64. http://doi.org/10.2307/40285811

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners1. *Japanese Psychological Research*, *46*(4), 337–349. http://doi.org/10.1111/j.1468-5584.2004.00265.x

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social*, *70*(3), 614.

Behrens, G., & Green, S. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music*, *21*(1), 20–33.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. http://doi.org/10.3758/BRM.40.2.531

Bestelmeyer, P., Rouger, J., DeBruine, L., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, *117*(2), 217–223.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition &*, *19*(8), 1113–1139.

Buchanan, T. W. T., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study, *9*(3), 227–238. http://doi.org/10.1016/S0926-6410(99)00060-9

Charlesworth, W. R., & Kreutzer, M. A. (1973). “Facial expression of infants and children.” In P. Ekman (Ed.), *Darwin and Facial Expression: a Century of Research in Review* (pp. 91–168). New York, NY: Academic Press.

Cottrell, S., & Mantzourani, E. (2006). The Clarinet and its players in Eastern Europe and Greece. In R. Heaton (Ed.), *The Versatile Clarinet* (pp. 31–46). New York, NY: Routledge.

Dailey, M., Cottrell, G. W., & Reilly, J. (2001). *California Facial Expressions,CAFE: Unpublished Digital Images*. San Diego, CA: University of California, Computer Science and Engineering Department.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*(7), 1019–1028. http://doi.org/10.3758/BF03194831

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), 1–10. http://doi.org/10.1016/S0010-0277(00)00136-0

Dellacherie, D., Hasboun, D., Baulac, M., Belin, P., & Samson, S. (2011). Impaired recognition of fear in voices and reduced anxiety after unilateral temporal lobe resection. *Neuropsychologia*, *49*(4), 618–629.

Dunlap, W., & Cornwell, J. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, *29*(1), 115–126.

Ekman, P. (1982). *Emotion in the Human Face*. Cambridge, MA: Cambridge University Press.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: Investigator’s Guide*. Palo Alto: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, C. J. (2002). *Facial Action Coding System Investigator’s Guide*. Salt Lake, UT: A Human Face.

Eschrich, S., Münte, T. F., Altenmüller, E. O., Panksepp, J., Eich, E., Metcalfe, J., … Halpern, A. (2008). Unforgettable film music: The role of emotion in episodic long-term memory for music. *BMC Neuroscience*, *9*(1), 48. http://doi.org/10.1186/1471-2202-9-48

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations, *36*(2), 480–487. http://doi.org/10.1016/j.neuroimage.2007.02.043

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts, *17*(3), 335–341. http://doi.org/10.3758/PBR.17.3.335

Flom, R., Gentile, D. A., & Pick, A. D. (2008). Infants' discrimination of happy and sad music. *Infant Behavior and Development*, *31*(4), 716–728. http://doi.org/10.1016/j.infbeh.2008.04.004

Frank, M., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social*, *80*(1), 75.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., … Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, *19*(7), 573–576. http://doi.org/10.1016/j.cub.2009.02.058

Gabrielsson, A., & Juslin, P. N. (1996). Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music*, *24*(1), 68–91. http://doi.org/10.1177/0305735696241007

Gabrielsson, A., & Juslin, P. N. (2003). *Emotional expression in music*. Oxford University Press.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244. http://doi.org/10.1016/j.neuropsychologia.2006.07.012

Gosselin, N., Peretz, I., Noulhiane, M., & Hasboun, D. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, *128*(3), 628–640.

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *http://dx.doi.org/10.1080/17470210902765957*, *62*(11), 2141–2155.

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception*, *23*(4), 319–330. http://doi.org/10.1525/mp.2006.23.4.319

Juslin, P., & Laukka, P. (2003). Emotional expression in speech and music. *Annals of the New York Academy of*, *1000*(1), 279–282.

Juslin, P. N. P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. http://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575. http://doi.org/10.1017/S0140525X08005293

Kantrowitz, J., Leitman, D., Lehrfeld, J., & Laukka, P. (2011). Reduction in tonal discriminations predicts receptive emotion processing deficits in schizophrenia and schizoaffective disorder. *Schizophrenia Bulletin*, *39*(1), 86–93.

Koeda, M., Belin, P., Hama, T., Masuda, T., & Matsuura, M. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology*, *4*(105).

Kotz, S., Meyer, M., Alter, K., Besson, M., & Cramon, D. von. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain and Language*, *86*(3), 366–76.

Lang, P. J., Öhman, A., & Vaitl, D. (1988). *The International Affective Picture System*. Gainesville, FL: University of Florida, Center for Research in Psychology.

Lima, C. F., & Castro, S. L. (2011). Emotion recognition in music changes across the adult life span. *Cognition and Emotion*, *25*(4), 585–98.

Lima, C., Garrett, C., & Castro, S. (2013). Not all sounds sound the same: Parkinson's disease affects differently emotion processing in music and in speech prosody. *Journal of Clinical and Experimental*, *35*(4), 373–92.

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410–1421.

Monrad-Krohn, G. H. (1963). "The third element of speech: prosody and its disorders." In L. Halpern (Ed.), *Problems of Dynamic Neurology* (pp. 101–107). Jerusalem: Hebrew University Press.

Omar, R., Henley, S., Bartlett, J., & Hailstone, J. (2011). The structural neuroanatomy of music emotion recognition: evidence from frontotemporal lobar degeneration. *Neuroimage*, *56*(3), 1814–21.

Pell, M. (2006). Judging emotion and attitudes from prosody following brain damage. *Progress in Brain Research*, *156*(303–317).

Peretz, I. (2010). Towards a Neurobiology of Musical Emotions. In J. Juslin, P; Sloboda (Ed.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 99–126). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199230143.003.0005

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., & B, J. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron*, *17*(33), 185–91.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. http://doi.org/10.1016/S0010-0277(98)00043-2

Peretz, I., & Zatorre, R. (2005). Brain organization for music processing. *Annu. Rev. Psychol.*, *56*, 89–114.

Roy, M., Mailhot, J.-P., Gosselin, N., Paquette, S., & Peretz, I. (2009). Modulation of the startle reflex by pleasant and unpleasant music. *International Journal of Psychophysiology*, *71*(1), 37–42. http://doi.org/10.1016/j.ijpsycho.2008.07.010

Salimpoor, V. V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature*, *14*(2), 257–262. http://doi.org/10.1038/nn.2726

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408–2412. http://doi.org/10.1073/pnas.0908239106

Scherer, K. (1994). Affect bursts: Essays on emotion theory. In S. H. M. van Goozen, N. E. van de Poll, & Sergeant J. A. (Eds.), *Emotions: Essays on emotion theory* (pp. 161–193). Hillsdale, NJ: Lawrence Erlbaum.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. http://doi.org/10.1037/0033-2909.99.2.143

Schirmer, A., Striano, T., & Friederici, A. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635–639.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication*. 40(1), 99-116.

Terwogt, M. M., & Van Grinsven, F. (1991). Musical Expression of Moodstates. *Psychology of Music*, *19*(2), 99–109. http://doi.org/10.1177/0305735691192001

Thompson, W., & Marin, M. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the*, *109*(46), 19027–32.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. http://doi.org/10.1037/1528-3542.8.4.494

Article 1: Tableaux et Figures

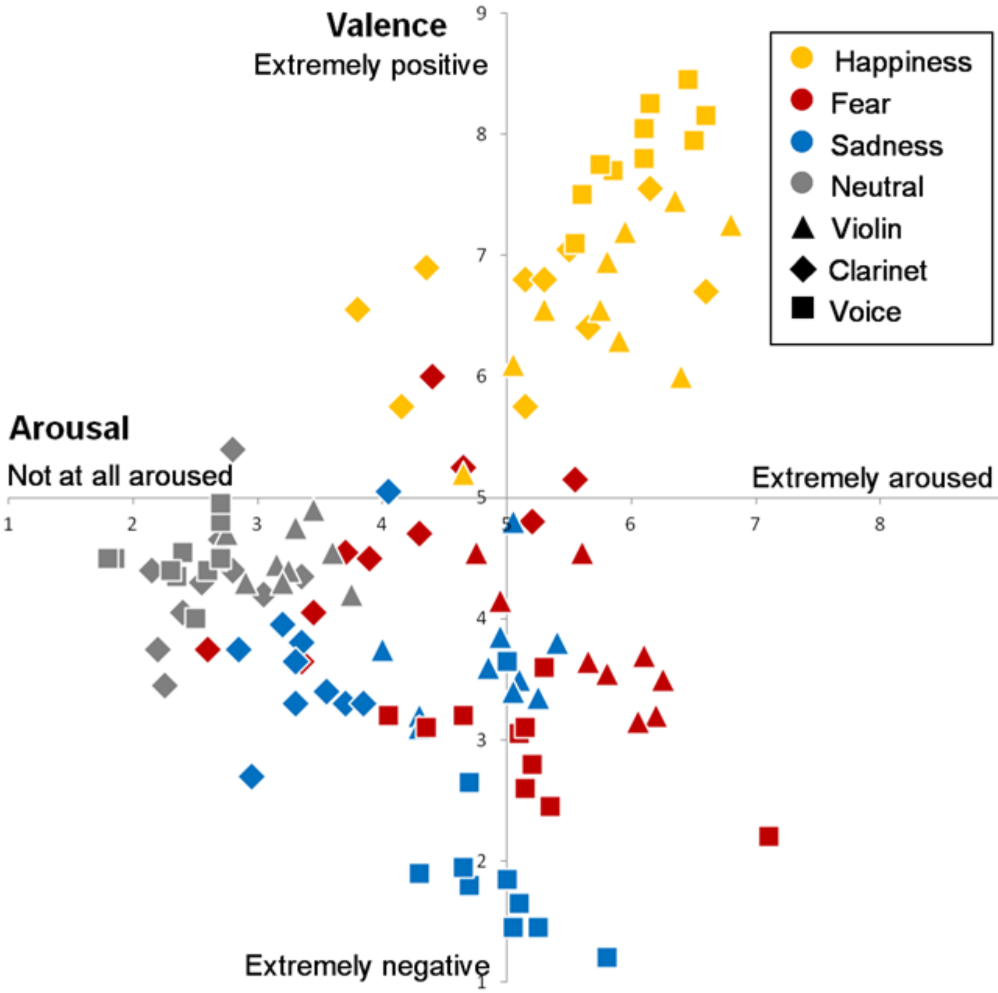
Tableau I. A1: Pourcentage d'identification correcte pour les MEB et MAV

Confusion matrix of emotion recognition for the MEB and MAV. Each row represents the percentage of choices for each emotion in each timbre. Percentage of correct recognition is presented in bold, (SE).

	Intended emotion	Forced Choice Happiness	Fear	Sadness	Neutral
Violin MEB	Happiness	**76.0 (3.1)**	13.0	4.0	7.0
	Fear	6.5	**88.0 (3.9)**	0.5	5.0
	Sadness	4.0	5.5	**88.0 (3.4)**	2.5
	Neutral	2.0	2.5	5.5	**90.0 (2.9)**
Clarinet MEB	Happiness	**92.0 (2.0)**	2.0	0.5	5.50
	Fear	15.0	**47.5 (4.6)**	13.0	24.5
	Sadness	3.0	9.5	**80.5 (4.1)**	7.0
	Neutral	2.0	3.5	13.0	**81.5 (4.2)**
Voice MAV	Happiness	**98.5 (1.1)**	0.0	1.5	0.0
	Fear	1.5	**93.0 (2.2)**	2.0	3.5
	Sadness	3.5	.5	**96.0 (1.5)**	0.0
	Neutral	1.0	7.5	0.0	**91.5 (4.5)**

Figure 1. A1: Jugements de valence et d'arousal pour chaque stimulus par timbre

Valence and arousal ratings for each stimulus played either on violin, clarinet, or voice as a function of the emotional intention.



Article 1: Supplementary Material

Tableau II. A1: Matériel supp. 1: Caractéristiques des stimuli au violon

S1. Characteristics of violin stimuli. For each stimulus, the duration (second), fundamental frequency (Hertz), ratings and percentage of correct recognition average by stimuli. At the bottom of the tables averages by emotions are presented (Standard deviation).

Name	Timbre	Type of Stimuli	Duration	meanF0	minF0	maxF0	Arousal	Valence	% correct recognition
V1_HAPPINESS(IMI)	Violin	IMITATION	1,30	449,2 (18,31)	438,50	499,68	5,90 (2,22)	6,30 (1,63)	70
V2_HAPPINESS(IMI)	Violin	IMITATION	1,15	232,76 (144,74)	80,25	609,66	6,40 (1,73)	6,00 (1,21)	40
V3_HAPPINESS(IMI)	Violin	IMITATION	1,58	524,36 (37,91)	478,94	597,21	4,65 (1,69)	5,20 (1,74)	45
V4_HAPPINESS(IMP)	Violin	IMPROVISATION	1,36	108,65 (6,77)	77,28	114,82	5,30 (2,20)	6,55 (1,32)	70
V5_HAPPINESS(IMP)	Violin	IMPROVISATION	1,46	504,24 (65,04)	377,64	586,29	5,80 (1,82)	6,95 (1,19)	95
V6_HAPPINESS(IMP)	Violin	IMPROVISATION	1,76	284,35 (125,44)	144,09	492,90	6,80 (1,54)	7,25 (1,21)	90
V7_HAPPINESS(IMP)	Violin	IMPROVISATION	1,75	390,84 (107,70)	203,33	591,18	6,35 (2,03)	7,45 (1,23)	100
V8_HAPPINESS(IMP)	Violin	IMPROVISATION	0,94	422,18 (62,05)	328,29	501,89	5,95 (1,82)	7,20 (1,36)	100
V9_HAPPINESS(IMP)	Violin	IMPROVISATION	1,08	505,82 (134,30)	256,16	624,70	5,05 (2,52)	6,10 (1,74)	75
V10_HAPPINESS(IMP)	Violin	IMPROVISATION	1,93	505,05 (88,34)	376,99	621,19	5,75 (1,92)	6,55 (1,70)	75
V1_FEAR(IMI)	Violin	IMITATION	0,77	470,18 (49,05)	386,70	615,78	4,75 (2,55)	4,55 (1,57)	75
V2_FEAR(IMI)	Violin	IMITATION	0,99	429,99 (27,01)	377,94	619,12	4,95 (2,67)	4,15 (1,35)	75
V3_FEAR(IMI)	Violin	IMITATION	1,00	260,65 (81,44)	164,32	422,80	5,60 (2,93)	4,55 (2,16)	85
V4_FEAR(IMP)	Violin	IMPROVISATION	1,04	347,36 (115,94)	143,75	588,24	5,65 (2,76)	3,65 (1,84)	90
V5_FEAR(IMP)	Violin	IMPROVISATION	0,94	522,37 (26,99)	434,74	540,58	6,10 (2,92)	3,70 (1,92)	95
V6_FEAR(IMP)	Violin	IMPROVISATION	0,94	533,08 (31,13)	423,92	589,80	6,20 (3,05)	3,20 (1,79)	90
V7_FEAR(IMP)	Violin	IMPROVISATION	1,18	321,21 (150,95)	128,28	602,28	5,80 (2.71)	3,55 (1,90)	100
V8_FEAR(IMP)	Violin	IMPROVISATION	1,07	378,83 (136,28)	157,79	493,98	6,25 (2,47)	3,50 (1,76)	85
V9_FEAR(IMP)	Violin	IMPROVISATION	1,16	432,13 (97,07)	297,65	529,56	6,10 (2,79)	3,70 (1,84)	95
V10_FEAR(IMP)	Violin	IMPROVISATION	1,01	514,99 (22,69)	482,15	571,17	6,05 (2,68)	3,15 (1,98)	90
V1_SADNESS(IMI)	Violin	IMITATION	2,20	397,43 (104,06)	263,12	635,39	4,00 (1,75)	3,75 (1,33)	75
V2_SADNESS(IMI)	Violin	IMITATION	2,24	249,38 (13,45)	221,38	266,90	4,30 (2,00)	3,20 (1,47)	85
V3_SADNESS(IMI)	Violin	IMITATION	2,94	412,82 (80,98)	263,35	636,52	4,30 (1,78)	3,10 (1,33)	85
V4_SADNESS(IMP)	Violin	IMPROVISATION	2,60	459,12 (13,48)	424,60	478,31	5,40 (2,09)	3,80 (1,79)	90
V5_SADNESS(IMP)	Violin	IMPROVISATION	3,10	216,87 (13,72)	193,15	233,01	4,85 (2,72)	3,60 (1,60)	95
V6_SADNESS(IMP)	Violin	IMPROVISATION	3,09	183,45 (24,38)	111,92	210,10	5,25 (2,20)	3,35 (1,60)	90
V7_SADNESS(IMP)	Violin	IMPROVISATION	2,38	539,85 (105,13)	274,25	632,01	4,95 (2,04)	3,85 (1,60)	95
V8_SADNESS(IMP)	Violin	IMPROVISATION	2,14	240,09 (28,54)	195,19	265,84	5,05 (1,82)	4,80 (1,11)	85
V9_SADNESS(IMP)	Violin	IMPROVISATION	2,24	361,59 (79,65)	279,06	471,46	5,05 (1,39)	3,40 (1,39)	95
V10_SADNESS(IMP)	Violin	IMPROVISATION	1,76	420,29 (12,70)	401,91	440,33	5,10 (1,94)	3,50 (1,47)	85
V1_NEUTRAL(IMI)	Violin	IMITATION	1,45	460,06 (1,14)	454,28	461,46	3,75 (2,10)	4,20 (1,40)	85
V2_NEUTRAL(IMI)	Violin	IMITATION	1,32	468,23 (0,67)	462,20	469,00	3,15 (1,95)	4,45 (1,00)	90
V3_NEUTRAL(IMI)	Violin	IMITATION	1,20	459,86 (1,15)	458,49	469,43	3,25 (2.00)	4,40 (1,35)	95
V4_NEUTRAL(IMP)	Violin	IMPROVISATION	1,15	262,21 (0,53)	258,32	262,82	2,65 (2,08)	4,50 (1,40)	95
V5_NEUTRAL(IMP)	Violin	IMPROVISATION	1,24	373,54 (0,99)	368,33	375,35	3,45 (2,04)	4,90 (1,33)	90
V6_NEUTRAL(IMP)	Violin	IMPROVISATION	1,30	494,58 (4,60)	458,66	496,91	3,60 (2,23)	4,55 (1,19)	90
V7_NEUTRAL(IMP)	Violin	IMPROVISATION	0,83	371,23 (1,04)	367,19	372,50	3,30 (2,08)	4,75 (1,33)	85
V8_NEUTRAL(IMP)	Violin	IMPROVISATION	1,25	314,54 (32,01)	279,00	565,97	3,20 (1,70)	4,30 (1,38)	90
V9_NEUTRAL(IMP)	Violin	IMPROVISATION	1,10	381,64 (0,79)	378,26	383,14	2,75 (1,97)	4,70 (1,13)	90
V10_NEUTRAL(IMP)	Violin	IMPROVISATION	1,16	294,7 (1,73)	280,47	296,01	2,90 (2,13)	4,30 (1,38)	90
Mean_HAPPINESS	Violin	All	1,43 (0,32)	392,74 (140,34)	276,15 (145,76)	523,95 (152,81)	5,80 (0,65)	6,56 (0,69)	
Mean_FEAR	Violin	All	1,01 (0,12)	421,08 (92,60)	299,72 (138,37)	557,33 (62,02)	5,75 (0,52)	3,77 (0,50)	
Mean_SADNESS	Violin	All	2,47 (0,45)	348,09 (118,72)	262,79 (94,18)	426,99 (173,34)	4,83 (0,46)	3,64 (0,48)	
Mean_NEUTRAL	Violin	All	1,20 (0,16)	388,06 (80,70)	376,52 (81,46)	415,26 (93,84)	3,20 (0,36)	4,51 (0,22)	

Tableau III.   A1: Matériel supp. 2: Caractéristiques des stimuli à la clarinette

S2. Characteristics of clarinet stimuli. For each stimulus, the duration (second), fundamental frequency (Hertz), ratings and percentage of correct recognition average by stimuli. At the bottom of the tables averages by emotions are presented (Standard deviation).

Name	Timbre	Type of Stimuli	Duration	meanF0	minF0	maxF0	Arousal	Valence	% correct recognition
C1_HAPPINESS(IMI)	Clarinet	IMITATION	1,16	403,36 (38,14)	333,03	449,54	4,35 (1,95)	6,90 (1,41)	90
C2_HAPPINESS(IMI)	Clarinet	IMITATION	1,75	404,82 (44,41)	283,26	621,35	4,15 (1,95)	5,75 (1,48)	80
C3_HAPPINESS(IMI)	Clarinet	IMITATION	1,58	342,08 (25,76)	261,86	357,15	5,15 (1,84)	5,75 (1,65)	65
C4_HAPPINESS(IMP)	Clarinet	IMPROVISATION	2,22	501,76 (72,10)	325,74	594,88	6,60 (1,67)	6,70 (1,92)	100
C5_HAPPINESS(IMP)	Clarinet	IMPROVISATION	1,39	339,34 (87,92)	234,07	603,00	6,15 (2,11)	7,55 (1,28)	100
C6_HAPPINESS(IMP)	Clarinet	IMPROVISATION	1,64	536,18 (100,89)	269,36	626,54	5,30 (2,18)	6,80 (1,01)	100
C7_HAPPINESS(IMP)	Clarinet	IMPROVISATION	1,05	387,18 (65,21)	297,19	475,32	3,80 (1,96)	6,55 (1,50)	95
C8_HAPPINESS(IMP)	Clarinet	IMPROVISATION	1,60	452,93 (55,44)	273,08	630,49	5,65 (2,11)	6,40 (1,73)	100
C9_HAPPINESS(IMP)	Clarinet	IMPROVISATION	1,60	525,05 (60,91)	383,44	616,38	5,15 (2,16)	6,80 (1,91)	95
C10_HAPPINESS(IMP)	Clarinet	IMPROVISATION	1,63	227,62 (28,71)	173,12	262,38	5,50 (2,35)	7,05 (1,79)	95
C1_FEAR(IMI)	Clarinet	IMITATION	0,99	382,19 (107,56)	172,56	554,69	5,20 (2,40)	4,80 (2,02)	65
C2_FEAR(IMI)	Clarinet	IMITATION	0,93	357,62 (92,91)	276,54	632,95	3,45 (1,93)	4,05 (1,05)	45
C3_FEAR(IMI)	Clarinet	IMITATION	1,10	368,02 (93,98)	183,24	621,98	5,55 (2,50)	5,15 (1,73)	60
C4_FEAR(IMP)	Clarinet	IMPROVISATION	0,75	411,71 (12,42)	388,31	420,61	3,70 (2,13)	4,55 (1,00)	40
C5_FEAR(IMP)	Clarinet	IMPROVISATION	1,07	164,44 (8,60)	152,84	175,43	3,35 (1,57)	3,65 (1,27)	25
C6_FEAR(IMP)	Clarinet	IMPROVISATION	0,98	552,66 (11,33)	524,26	560,95	3,90 (2,02)	4,50 (1,36)	55
C7_FEAR(IMP)	Clarinet	IMPROVISATION	1,00	183,67 (5,02)	169,02	190,38	2,60 (1,57)	3,75 (0,85)	40
C8_FEAR(IMP)	Clarinet	IMPROVISATION	0,55	474,71 (37,84)	397,40	501,04	4,30 (1,92)	4,70 (1,84)	50
C9_FEAR(IMP)	Clarinet	IMPROVISATION	0,40	419,39 (87,86)	277,50	517,24	4,65 (2,39)	5,25 (1,83)	65
C10_FEAR(IMP)	Clarinet	IMPROVISATION	0,49	497,73 (53,55)	406,85	602,31	4,40 (2,50)	6,00 (1,65)	30
C1_SADNESS(IMI)	Clarinet	IMITATION	3,03	328,78 (15,23)	308,27	352,83	3,70 (1,89)	3,30 (1,38)	85
C2_SADNESS(IMI)	Clarinet	IMITATION	2,10	336,92 (16,28)	302,92	352,14	3,35 (1,60)	3,80 (0,95)	85
C3_SADNESS(IMI)	Clarinet	IMITATION	2,61	345,95 (75,60)	145,34	396,02	4,05 (1,76)	5,05 (1,50)	70
C4_SADNESS(IMP)	Clarinet	IMPROVISATION	2,10	361,8 (10,09)	351,72	376,58	2,85 (1,73)	3,75 (1,33)	80
C5_SADNESS(IMP)	Clarinet	IMPROVISATION	1,73	390,5 (28,33)	352,25	422,42	3,30 (2,15)	3,65 (1,14)	80
C6_SADNESS(IMP)	Clarinet	IMPROVISATION	2,38	193,65 (13,53)	176,95	211,97	3,30 (2,25)	3,30 (0,98)	75
C7_SADNESS(IMP)	Clarinet	IMPROVISATION	2,57	260,52 (12,17)	245,61	280,38	3,55 (1,93)	3,40 (1,50)	75
C8_SADNESS(IMP)	Clarinet	IMPROVISATION	3,02	303,16 (9,19)	290,02	312,96	2,95 (1,90)	2,70 (1,26)	80
C9_SADNESS(IMP)	Clarinet	IMPROVISATION	2,74	210,55 (12,58)	194,23	222,35	3,85 (2,58)	3,30 (1,17)	85
C10_SADNESS(IMP)	Clarinet	IMPROVISATION	3,11	243,62 (13,63)	232,94	262,03	3,20 (2,28)	3,95 (1,79)	90
C1_NEUTRAL(IMI)	Clarinet	IMITATION	1,27	233,77 (1,00)	232,41	237,55	2,55 (1,28)	4,30 (1,17)	85
C2_NEUTRAL(IMI)	Clarinet	IMITATION	1,59	233,61 (0,86)	231,84	235,93	2,40 (1,79)	4,05 (1,10)	80
C3_NEUTRAL(IMI)	Clarinet	IMITATION	1,77	395,17 (1,03)	393,55	396,65	3,05 (1,57)	4,20 (1,32)	85
C4_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,14	442,83 (0,33)	441,01	443,44	2,80 (1,67)	5,40 (1,60)	90
C5_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,62	418,51 (1,71)	414,75	422,66	2,70 (1,53)	4,65 (1,35)	80
C6_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,74	525,38 (2,62)	513,00	529,74	3,35 (1,90)	4,35 (1,39)	80
C7_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,43	177,36 (0,28)	177,08	178,48	2,20 (1,54)	3,75 (1,12)	75
C8_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,44	374,05 (1,00)	371,24	376,60	2,15 (1,57)	4,40 (1,27)	75
C9_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,70	354,39 (0,33)	353,86	355,07	2,80 (1,67)	4,40 (1,05)	80
C10_NEUTRAL(IMP)	Clarinet	IMPROVISATION	1,60	148,17 (0,83)	143,29	151,69	2,25 (1,80)	3,45 (1,19)	85
Mean_HAPPINESS	Clarinet	All	1,56 (0,32)	412,03 (95,96)	283,41 (57,56)	523,70 (131,37)	5,18 (0,88)	6,63 (0,55)	
Mean_FEAR	Clarinet	All	0,83 (0,26)	381,21 (125,10)	294,852 (128,44)	477,76 (167,53)	4,11 (0,89)	4,64 (0,72)	
Mean_SADNESS	Clarinet	All	2,54 (0,46)	297,55 (66,93)	260,02 (72,50)	318,97 (72,94)	3,41 (0,38)	3,62 (0,62)	
Mean_NEUTRAL	Clarinet	All	1,53 (0,21)	330,33 (124,91)	327,20 (123,21)	332,78 (125,02)	2,625 (0,39)	4,30 (0,51)	

Tableau IV. A1: Matériel supp. 3: Caractéristiques des stimuli vocaux

S3. Characteristics of vocal stimuli. For each stimulus, the duration (second), fundamental frequency (Hertz), ratings and percentage of correct recognition average by stimuli. At the bottom of the tables averages by emotions are presented (Standard deviation).

Name	Timbre	Type of Stimuli	Duration	meanF0	minF0	maxF0	Arousal	Valence	% correct recognition
42_happiness	Voice	MAV	1,44	169,02 (24,77)	124,21	227,17	5,85 (1,76)	7,70 (1,30)	95
45_happiness	Voice	MAV	1,56	369,6 (50,05)	318,34	505,05	5,75 (2,02)	7,75 (0,91)	100
46_happiness	Voice	MAV	1,01	392,04 (141,09)	212,15	570,16	6,10 (2,27)	8,05 (0,89)	100
53_happiness	Voice	MAV	0,96	257,19 (51,71)	143,85	324,21	5,60 (2,09)	7,50 (1,10)	100
55_happiness	Voice	MAV	1,10	277,99 (123,03)	151,44	555,07	5,55 (2,09)	7,10 (1,25)	95
58_happiness	Voice	MAV	1,05	324,28 (47,82)	234,91	416,28	6,10 (1,74)	7,80 (1,11)	95
59_happiness	Voice	MAV	1,83	461,95 (84,03)	179,44	595,11	6,45 (2,54)	8,45 (1,00)	100
6_happiness	Voice	MAV	1,74	235,29 (56,46)	123,94	323,43	6,50 (1,79)	7,95 (1,05)	100
60_happiness	Voice	MAV	1,16	381,75 (154,16)	107,67	603,21	6,60 (2,04)	8,15 (0,81)	100
61_happiness	Voice	MAV	2,60	228,72 (137,52)	121,07	587,97	6,15 (2,30)	8,25 (0,85)	100
42_fear	Voice	MAV	0,41	270,89 (49,12)	149,77	312,23	4,65 (2,83)	3,20 (1,36)	95
45_fear	Voice	MAV	0,63	327,68 (53,34)	269,38	591,66	5,20 (2,59)	2,80 (1,64)	85
46_fear	Voice	MAV	0,81	425,26 (130,24)	177,43	600,67	7,10 (2,47)	2,20 (1,77)	100
53_fear	Voice	MAV	0,84	445,41 (40,42)	333,56	476,77	5,10 (2,20)	3,05 (1,36)	85
55_fear	Voice	MAV	0,61	278,76 (23,33)	205,25	301,94	4,05 (1,93)	3,20 (1,32)	85
58_fear	Voice	MAV	0,49	422,89 (19,02)	389,20	451,69	5,15 (2,35)	3,10 (1,55)	100
59_fear	Voice	MAV	0,72	315,88 (40,54)	208,69	359,55	4,35 (2,16)	3,10 (1,68)	95
6_fear	Voice	MAV	0,76	259,13 (76,39)	128,62	332,28	5,35 (2,62)	2,45 (1,39)	100
60_fear	Voice	MAV	0,44	507,83 (78,48)	316,74	578,47	5,30 (2,45)	3,60 (1,64)	90
61_fear	Voice	MAV	0,32	398,3 (50,32)	319,06	484,99	5,15 (2,62)	2,60 (1,64)	95
42_sadness	Voice	MAV	1,67	208,52 (26,33)	148,10	261,35	5,00 (2,18)	3,65 (2,06)	80
45_sadness	Voice	MAV	1,78	428,51 (97,92)	255,98	631,99	5,10 (2,63)	1,65 (1,09)	95
46_sadness	Voice	MAV	1,96	392,94 (92,73)	126,11	610,80	5,05 (2,78)	1,45 (0,94)	100
53_sadness	Voice	MAV	2,88	312,03 (60,73)	141,18	521,65	4,70 (2,52)	1,80 (0,95)	100
55_sadness	Voice	MAV	1,83	262,86 (92,25)	120,93	577,01	4,70 (2,32)	2,65 (1,95)	95
58_sadness	Voice	MAV	1,42	355,97 (92,12)	186,65	543,73	4,65 (2,48)	1,95 (1,54)	100
59_sadness	Voice	MAV	3,60	374,17 (115,70)	198,59	580,44	5,25 (2,61)	1,45 (1,05)	90
6_sadness	Voice	MAV	1,64	287,74 (82,24)	122,98	586,93	5,00 (2,32)	1,85 (1,09)	100
60_sadness	Voice	MAV	2,38	359,15 (76,96)	233,60	559,22	5,80 (2,71)	1,20 (0,89)	100
61_sadness	Voice	MAV	2,44	207,03 (116,97)	108,37	584,68	4,30 (2,64)	1,90 (1,12)	100
42_neutral	Voice	MAV	1,31	123,47 (73,27)	104,75	594,16	2,40 (1,39)	4,55 (1,10)	95
45_neutral	Voice	MAV	0,99	228,52 (2,57)	225,11	239,99	2,70 (2,20)	4,80 (1,15)	90
46_neutral	Voice	MAV	0,24	261,46 (4,78)	254,82	269,06	2,35 (1,66)	4,35 (1,39)	85
53_neutral	Voice	MAV	0,95	190,36 (1,55)	184,63	194,89	2,60 (1,79)	4,40 (0,82)	95
55_neutral	Voice	MAV	1,24	109,43 (1,80)	104,04	114,25	1,85 (1,14)	4,50 (1,36)	95
58_neutral	Voice	MAV	0,51	220,62 (51,72)	184,28	461,53	2,50 (1,28)	4,00 (1,30)	75
59_neutral	Voice	MAV	0,65	143,4 (1,68)	138,81	147,15	2,30 (2,05)	4,40 (1,39)	95
6_neutral	Voice	MAV	0,90	112,14 (4,75)	86,28	115,45	2,70 (2,08)	4,50 (1,00)	95
60_neutral	Voice	MAV	1,60	216,97 (21,92)	206,75	379,67	2,70 (1,22)	4,95 (1,23)	90
61_neutral	Voice	MAV	1,86	94,56 (1,19)	88,58	98,02	1,80 (1,24)	4,50 (1,05)	100
Mean_HAPPINESS	Voice	All	1,45 (0,51)	309,78 (91,04)	171,70 (66,32)	470,77 (137,60)	6,07 (0,37)	7,87 (0,39)	
Mean_FEAR	Voice	All	0,60 (0,18)	365,20 (85,81)	249,77 (88,01)	449,03 (117,45)	5,14 (0,81)	2,93 (0,42)	
Mean_SADNESS	Voice	All	2,16 (0,85)	318,89 (76,31)	164,25 (51,48)	545,78 (104,73)	4,96 (0,41)	1,96 (0,71)	
Mean_NEUTRAL	Voice	All	1,02 (0,49)	170,09 (60,14)	157,81 (61,22)	261,42 (167,47)	2,39 (0,33)	4,50 (0,26)	

Figure 2. A1: Matériel supp. 4: Fréquences spectrales et formes d'ondes (violon).

S4. Spectral frequency and temporal waveform of the violin stimuli.

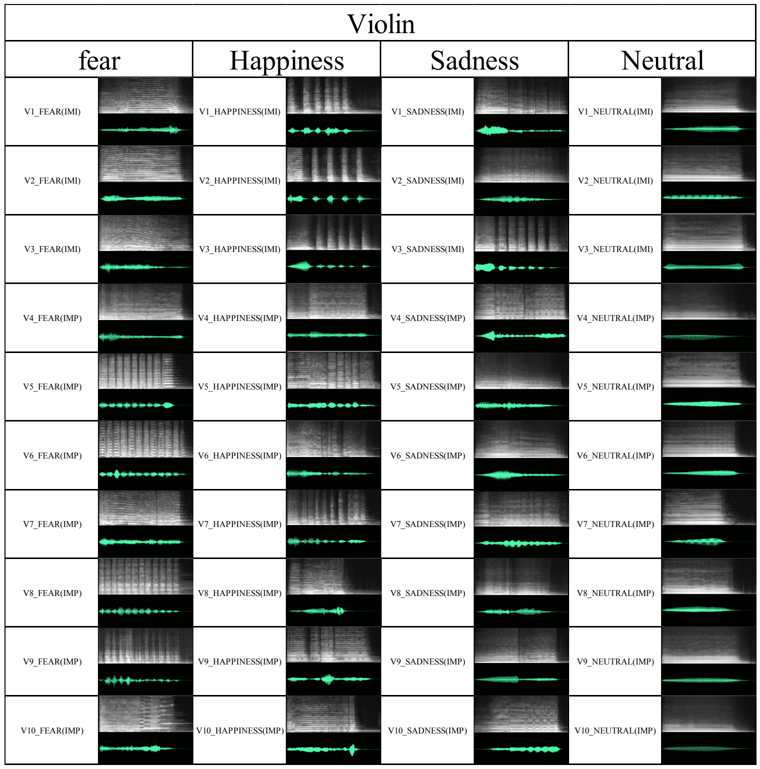
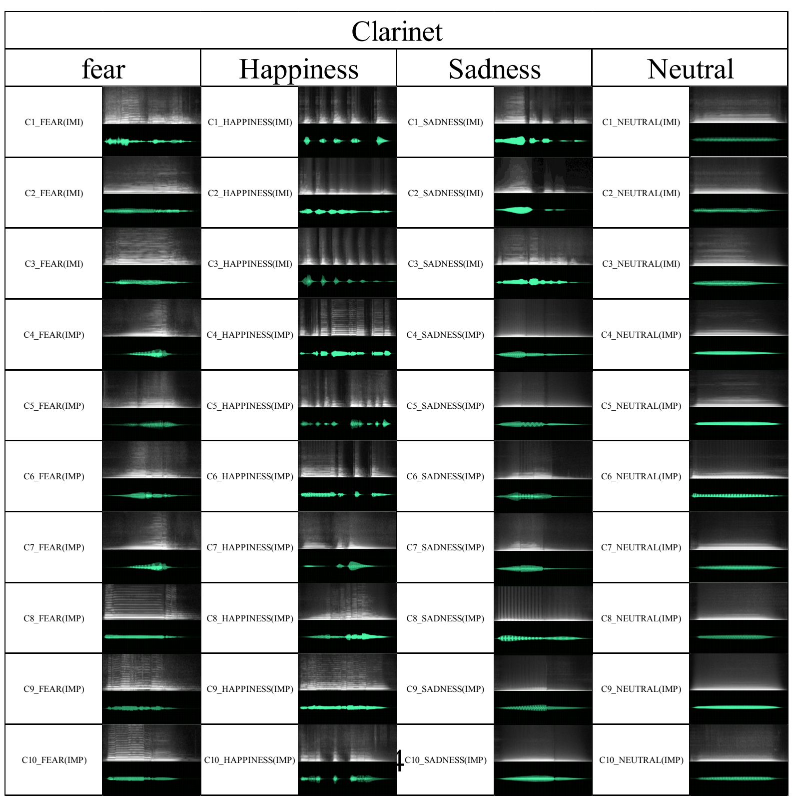


Figure 3. A1: Matériel supp. 5: Fréquences spectrales et formes d'ondes (clarinette)

S5. Spectral frequency and temporal waveform of the clarinet stimuli.



S6. For the spectral frequency waveforms of the vocal stimuli, see:

Belin, P., Fillion-Bilodeau, S., Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. Behavior Research Methods, 40(2), 531-539.

Tableau V. A1: Matériel supp. 7: Pourcentage d'identification (stimuli improvisés)

S7. Percentage of correct identification of the improvised stimuli as a function of timbre and intended emotion (SE).

	Violin	Clarinet	Voice
Happiness	86.4 (3.5)	97.9 (1.2)	98.5 (1.1)
Fear	92.1 (4.3)	43.6 (5.8)	93.0 (2.2)
Sadness	90.7 (3.5)	80.7 (4.8)	96.0 (1.5)
Neutral	90.0 (3.3)	80.7 (4.6)	91.5 (4.5)

Tableau VI. A1: Matériel supp. 8: Pourcentage d'identification (stimuli imités)

S8. Percentage of correct identification of the imitated stimuli as a function of timbre and intended emotion (SE).

	Violin	Clarinet
Fear	78.3 (6.1)	56.7 (6.4)
Happiness	51.7 (7.0)	78.3 (5.6)
Sadness	81.7 (6.2)	80.0 (4.5)
Neutral	90.0 (4.3)	83.3 (5.1)

A2: Musical and vocal emotions - Are they all the same?

Sébastien Paquette[1], Isabelle peretz[1], Pascal Belin[1,2,3]

[1]Department of Psychology, International Laboratory for Brain Music and Sound Research, Center for Research on Brain, Language and Music, University of Montreal, Montreal, QC, Canada

[2] Voice Neurocognition Laboratory, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

[3] Institut des Neurosciences de La Timone, Aix-Marseille Université, Marseille, France

Keywords: Music, Voice, Emotion, Cross-domain comparison, Auditory Cortex, Amygdala

Musical and vocal emotions - Are they all the same?

Abstract

In the last decades there has been growing evidence supporting the idea of common neural substrates for processing musical and vocal emotions. Both stimuli (timbre) activate similar cortical (Heschl's Gyrus, Planum temporale) and subcortical (amygdala/hippocampus) brain regions but with the occasional timbre-specific activity that may reflect a difference in processing for these stimuli. We hypothesize that the observed differences are not specific to timbre, but due to stimuli-specific factors such as emotional complexity and length of stimuli that can differ between sound categories.

Here, we circumvent stimulus comparability issues by using a novel set of similar short musical and vocal affect bursts that share common stimulus characteristics and depict basic emotional expressions such as fear, happiness, sadness and neutrality to compare brain activity between vocal and musical sounds with functional Magnetic Resonance Imaging.

Emotion-specific activity was found in the auditory cortex and in the limbic regions, while no activity specific to vocal or musical stimuli was observed; all emotions (happiness, sadness, fear) generated differentiable activity in the primary auditory cortex independently of their vocal or musical nature. Additionally, fear stimuli elicited more activity than their neutral counterpart in the amygdala/parahippocampal gyrus. Taken together, this observed emotion-specific activity in cortical and subcortical regions irrespective of timbre supports the idea of a common auditory emotional neural network for both music and voice.

Introduction

Affective sounds are important for social interactions because they structure and influence behaviors. Such sounds, like screams and laughter, are vocal bursts (vocalizations, non-verbal interjections; Scherer, 1994) that typically accompany intense emotions and convey immediate indications of a person's affective state. Similarly, music can effectively communicate emotions. Musical emotion can be recognized quickly, in as little as a quarter of a second, or the equivalent of one chord or a few notes (Bigand, Filipic, & Lalitte, 2005; Peretz et al., 1998).

Several studies (Curtis & Bharucha, 2010; Ilie & Thompson, 2006), including a meta-analysis by Juslin and Laukka (2003), have identified several similarities between vocal and musical expression regarding the emotion-specific patterns of acoustic cues (e.g. tempo, intensity) used to communicate emotions. There is also a growing body of evidence suggesting that the ability to recognize emotions from vocal expressions and from music develops early in life (Juslin & Laukka, 2003) and has similar developmental trajectories (Laukka & Juslin, 2007). It is therefore not surprising that many researchers have theorized about the existence of common neural substrates for processing musical and vocal emotions.

One possibility is that music uses emotional circuits that have evolved primarily for the processing of biologically important vocalizations (Peretz, 2010). The common acoustical characteristics shared by both domains would promote neural recycling/co-opting of circuits (Dehaene & Cohen, 2007). Frühholz and collaborators (2016) have recently proposed a unifying cognitive and neural network model outlining how the brain decodes emotional meaning from various sound types. Their neural model involves at its core both the amygdala

(AMYG) and the auditory cortex (AC). Thereby, they suggest that the AMYG functions as a relevance detector by responding to sound features that can immediately identify sound importance, such as the rising intensity of a warning signal (Bach, Schächinger, & Neuhoff, 2008). Their model gives the AC a complementary role, namely the further treatment of sound characteristics that develop over time by integrating acoustic information into larger time-scale structures (Zatorre & Belin, 2001). In this context, the ability to perform auditory emotion discrimination and/or recognition would result from the constant communication between these two core regions (Frühholz et al., 2016).

There is a significant body of evidence highlighting the similarities in neural activation for non-emotional vocal and musical stimuli in the primary AC (see Schirmer, Fox, & Grandjean, 2012 for a meta-analysis). With regards to emotional processing, the majority of studies have reported the implication of the AMYG for vocal (Dellacherie et al., 2011; Fecteau et al., 2007; Scott et al., 1997; Sprengelmeyer et al., 1999) and musical emotional processing (Gosselin et al., 2005, 2007; Koelsch et al., 2013; Mitterschiffthaler et al., 2007; Salimpoor et al., 2011).

Generally, when directly compared, musical and vocal emotional stimuli seem to activate similar cortical (Heschl's Gyrus: HG; Planum Temporale: PT) and subcortical (AMYG/hippocampus and posterior insula) brain regions (Aubé et al., 2015; Escoffier et al., 2013). These same studies have also found timbre-specific activity within the auditory cortex. More specifically, voice processing was associated with higher activity in the middle/superior temporal cortex and in the superior temporal sulcus and music processing with higher activity in the PT/HG and Superior Temporal Gyrus (STG). These different activations could reflect timbre-specific processing, but could also be due, at least in part, to factors such as a difference in stimuli emotional complexity (varies over time) and/or length. It is important to note that the two previously mentioned neuroimaging studies (Aubé et al., 2015; Escoffier et al., 2013) have directly compared musical and vocal emotion perception using very different stimulus sets. The first (Escoffier et al., 2013) used long (> 10 seconds) musical and vocal stimuli and contrasted brain activation elicited by complex familiar music with that elicited by vocal utterances of vowels in a meaningless order. The second (Aubé et al., 2015) contrasted short excerpts of music written in the western tonal system (some on discrete pitch instrument) with vocalization stimuli without controlling for the segmental structure (phonological units; vowels, consonants) of the auditory stimuli.

Here, we circumvent stimulus comparability issues by using short and similar emotion exclamations (affect bursts) in music and voice to elicit similar (discrete) basic emotions and characterize cerebral activity for each domain (music or voice). To do so, we are using a two sets of validated highly comparable (ecological similarity and in emotional complexity) musical and vocal stimuli specifically designed to elicit basic emotions: The Montreal Affective Voices (MAV) and the Musical Emotional Bursts (MEB).

The MAV (Belin et al., 2008) are non-speech vocalizations from ten professional actors (e.g., screams, laughter) depicting basic emotions that are fundamental to human communication (Scherer, 1986). They all consist of short vocal interjections on the vowel /a/ expressing fear, happiness, sadness and neutrality (every actors expressed every emotion); they have minimal semantic information and minimal interaction with linguistic processes (Bestelmeyer et al., 2010). The MEB are the validated musical counterparts of the MAV and consist of short musical bursts (a few notes) from professiona musicians expressing fear, happiness, sadness and neutrality on continuous pitch instruments (clarinet and violin), which have a seamless progression between notes just like vocal stimuli. Both stimuli set have been previously compared on diverse emotional scale in Paquette and collaborators (2013).

Here, we used functional Magnetic Resonance Imaging (fMRI) to directly compare brain responses of young adults passively listening to basic emotions expressed through music and voices to see whether both types of sounds share an emotion network. We predicted to observe emotion-specific activity in the auditory cortex and the limbic regions and hypothesize that due to our careful stimuli selection, we will not observe cerebral activity that is timbre-specific but will observe common levels of activity for both musical and vocal emotions.

Method

Participants

Twenty participants (10 female, 19 right-handed; average Edinburgh Handedness Inventory score: 79.4) were recruited through the departmental subject mailing-list. They reported normal hearing and no neurological disorders. Participants were between 18 and 29 years old (m= 22.9) and had on average 5.6 years (0-18 years) of musical education. Participants all provided written informed consent prior to participation, in accordance with the Declaration of Helsinki. The local ethics committee at the University of Glasgow approved the protocol.

Stimuli

One hundred and twenty previously validated stimuli were used in the study: 40 (20 male, 20 female) short non-verbal emotional interjection (m= 1.3s) selected from the MAV

(Belin et al., 2008) and 80 (40 clarinet, 40 violin) short musical bursts (M=1.6s) selected from the MEB (Paquette et al., 2013). Ten stimuli from each timbre (voice, clarinet, and violin) expressed one of 4 emotions (fear, sadness, happiness, neutrality); their respective acoustical features are presented in Table 1. Since musical timbers differ in their ability to express specific emotions (Hailstone et al., 2009; Paquette et al., 2013) we used a larger number of musical stimuli, but analysed them (violin, clarinet) together as music for simplicity and better comparison with the current literature (total= 80 musical, 40 vocal stimuli). To make sure there was no difference global recognition accuracy between music and voice, participants performed a forced-choice emotion (fear, sadness, happiness, neutrality) identification task after the fMRI experiment. Participants mean recognition accuracy for each timbre (voice: 73.8%, clarinet: 70.9%, violin 82.8%) was significantly different ($p > 0.001$) than chance level (25%) and no significant difference ($t_{48} = 1.08$, $p = .28$) in emotion recognition accuracy was found between voice 73.8% (SD: 0.03) and music 76.8% (SD: 0.12).

[Insert Table 1 here]

Procedure

Participants performed a one-back task while listening to the affect bursts. Stimuli were presented in blocks composed of 40 randomly presented stimuli (ISI: 2-2.5 sec) from one of the three timbres (voice, clarinet, and violin) and included all 4 emotions (happy, sad, fear, neutral). Each block was presented 4 times in a random order and separated by 20 sec of silence. Stimuli were presented at a comfortable listening level over MRI-compatible electrostatic in-ear headphones (Sensimetrics Corporation, USA) using an M-Audio Audiophile 2496 soundcard. Participants were instructed to keep their eyes closed. To make sure participants were paying attention to the stimuli they were also instructed to press a button using their index finger on an MRI-compatible response pad (Lumitouch) every time they heard the same sound twice in a row (4 repetitions per block). Participants did not miss more than two repetitions in a row.

MRI Acquisition

Structural and blood oxygenation level dependent (BOLD) functional images were acquired on a 3T Tim Trio Scanner (Siemens) and a 32-channel head coil at the Centre for Cognitive Neuroimaging (CCNi) of the Institute of Neuroscience and Psychology, University of Glasgow. Functional images during the task were acquired using continuing fMRI scanning and a parallel-accelerated multi-echo Echo Planar Imaging (EPI) sequence (TE: 9.4/ 21.2/ 33.45/ 57). Volumes consisted of 32 axial slices (voxel size: 3 × 3 × 3 mm3; gap: 25%) and were acquired at a TR of 2.47 seconds. The acquisition ended with a T1-weighted anatomical scan (voxel size: 1 × 1 × 1 mm3).

MRI Preprocessing

Data pre-processing was performed using Statistical Parametric Mapping (SPM8; Wellcome Trust Centre for Neuroimaging) and analyses were performed using SPM12. All images and echoes were realigned to correct for head motion with the first volume as a reference. Echoes were then combined by weighted summation using the parallel-acquired inhomogeneity-desensitized (PAID) method. T1-weighted structural images were co-registered to the mean image created by the realignment procedure and were used for normalization of functional images onto the Montreal Neurological Institute (MNI) atlas using normalization parameters derived from segmentation of the anatomical image. Finally, each image was smoothed with an isotropic 8 mm full-width-at-half-maximum Gaussian kernel.

EPI time series were analyzed across the entire brain using the general linear model (GLM) as implemented in SPM12. At the subject level, the GLM included 9 conditions: two timbres (voice, music) x four emotions (happy, sad, fear, neutral) with the addition of repetitions and button presses as the ninth condition. Using the first eight conditions, contrasts of interest ‘‘each emotion/each timbre > baseline’’ were performed for each subject, and entered in standard group level analysis (2nd-level analysis; 2X4 ANOVA) comparing timbres (2) and emotions (4).

Results

Group level analysis revealed no main effect of timbre ($F_{1,152} = 23.80$, $p > 0.05$ Family-Wise Error (FWE)-corrected). However, a significant symmetrical main effect of emotion type ($F_{3,152} = 38.38$, $p < 0.001$ FWE; $F_{3,152} = 44.77$, $p < 0.001$ FWE) was obtained in the left and right AC (Centred in the STG; Figure 1; Blue). We also found a significant ($F_{3,152} = 12.50$, $p < 0.01$ FWE; $F_{3,152} = 13.54$, $p < 0.01$ FWE) main effect of emotion (Figure 1; Red) in the left and right AM/parahippocampal gyrus (PHG). Additionally, significant timbre x emotion interactions (Figure 1; Green) were also observed in the STG on both sides close to HG ($F_{3,152} = 11.50$, $p < 0.05$ FWE; $F_{3,152} = 11.61$, $p < 0.05$ FWE).

[Insert Figure 1 here]

In order to explore our main effects of emotions and explain our interaction, planned Bonferonni-corrected posthoc analyses contrasting each emotion and timbres were performed in SPM12 ($p < .05$ FWE – whole brain analysis) using previous main effects of emotion and interactions as masks.

Within the main effects of emotion in the left and right AC [Peak: -66 -19 -4; 57 5 -8] fear stimuli produced the most robust activation, which generated significantly higher BOLD responses (all $p < 0.005$ FWE; most significant peak within ±6x18x6 (x,y,z) voxels from the main effect peak) compared to other emotion types (see Figure 1a,b). Next were happy stimuli, which produced stronger BOLD activation than both neutral and sad stimuli (all $p < 0.005$ FWE; most significant peak within ±6x3x3 (x,y,z) voxels from the main effect peak). However, the activation produced by sad stimuli was not found to be statistically different than that produced by neutral stimuli ($p > 0.05$ FWE). Although they are not directly of interest when looking into a main effect of emotion, it is interesting to note that no contrasts comparing each emotion for each timbre reached significance ($p > 0.05$ FWE) after correction for multiple comparisons (except for those explained by our general emotion comparison).

The interaction observed in STG close to HG (Figure 2) can be explained by the fact that happy musical stimuli elicited more activity that it's vocal counterpart (Figure 1c, d). Although this observation was not supported by a significant difference, it resulted in other significant within timbre differences in the right and left STG. On the right, the interaction can be explained by the fact that vocal happy stimuli elicited significantly less activity than the vocal fear stimuli ($F_{1,152} = 34.68$, $p < 0.005$ FWE) while musical happy and fear stimuli were associated with similar levels of activity ($p > 0.05$ FWE). On the left side, the interaction can be explained by the fact that musical happy stimuli elicited much more activity than the musical sad stimuli ($F_{1,152} = 47.45$, $p < 0.005$ FWE) while the same comparisons did not reach significance for the vocal stimuli ($p > 0.05$ FWE). Despite the interaction, comparable effects of emotions as those observed for the AC were found in this region, with the only exception that in the left hemisphere the activity elicited by happy stimuli was not significantly different than that elicited by sad stimuli ($p > 0.05$ FWE).

As for the main effect observed in the left and right AM/PHG [peak: -33 2 -17; 18 -4 -17], similar effects to those observed in the AC for fear were found, with a few exceptions (Figure 1e, f). In the left AMYG/PHG, the activity elicited by fear stimuli was not significantly different than that elicited by sad stimuli. In both hemispheres, fear stimuli produced BOLD responses larger than those elicited by neutral stimuli (both $p < 0.05$ FWE), but not larger than those produced by happy stimuli (both $p > 0.05$ FWE). The vocal and musical fear contrast, as well as other contrasts comparing each emotion for each timbre, did not reach significance ($p > 0.05$ FWE), after correction for multiple comparisons.

[Insert Figure 2 here]

Discussion

The purpose of the present study was to examine the shared and distinct neural networks that underlie emotional processing of musical and vocal stimuli. Activation patterns elicited by vocal and musical emotions showed striking similarities as indicated by the lack of a significant main effect of Timbre (voice vs. music) for the group level (whole brain) analysis and held true for each emotion considered within the main effects of emotions. Within the left and right AC fear stimuli elicited more activity than all other emotions and happy stimuli were associated with more activity than the sad and neutral stimuli. Similarly, fear stimuli elicited significantly more activity than neutral stimuli in the AMYG/PHG. Of note, in the STG (close to HG), happy musical stimuli seem to be differentially processed than vocal happy stimuli.

The auditory region (STG, near HG) where a Timbre x Emotion interaction was found overlaps with the area of AC associated with the main effect of emotion. Within this area centered on the STG, the different emotional types presented generally created differential brain responses (Fear > Happiness > Sadness) irrespective of timbre. Higher activity (similar to fear) for happy musical stimuli modulated the main effect in a discrete area; as revealed by the interaction. Previous imaging studies that compared music and voice processing have typically found, in this area, more activation for musical sounds when compared to vocal sounds (Angulo-Perkins et al., 2014 [Peak:-50 -6 -4; 50 -4 -4]; Aubé et al., 2015 [-50 -2 -8; 50 4 12]). Although in the present study, this effect is specific to the happy musical stimuli, suggesting that this region could show a preference for music specific processing when the auditory stimulus conveys happiness. An alternative explanation for our results can be linked to a difference in valence between stimuli observed in the validation study (see figure1 of Paquette et al., 2013). Although some of the happy musical stimuli received the highest arousal ratings, they were still generally rated as less positive than laughter. Happy violin stimuli, when misclassified, were often identified as fear (a high arousal, low valence stimuli). This region of the STG could serve as an ‘‘arousal detector’’ and preferentially process highly arousing sounds as opposed to being a valence detector.

In the current study, the presented emotions (fear, happy, sad) were associated with different levels of brain activity in the AC, regardless of timbre and only sadness was not significantly different than neutral stimuli. As proposed by Aubé and collaborators (2015), in such cases, the absence of difference in brain activity between the sad and neutral stimuli could be driven by the arousal value of these stimuli. Indeed, fear, happy and sad/neutral stimuli can be differentiated on this aspect alone (Yik et al., 1999). Although plausible, this interpretation is not supported by the behavioral ratings obtained in the validation study in which 2/3 of the sad stimuli (violin, voice) were judged significantly more arousing than neutral ones (Paquette et al., 2013). An alternative explanation for the absence of difference in activity between sad and neutral stimuli would be from the acoustical point of view. As proposed by Frühholz and collaborators (2016) in their neurocognitive model for affective sounds processing, the AC could serve to decode more abstract sound features (previously observed with the MAV: Bestelmeyer, Maurage, Rouger, Latinus, & Belin, 2014). Acoustical features like brightness (energy above 3000 Hz) have been known to correlate with emotional judgments (Gosselin et al., 2015; Paquette, Ahmed, Peretz, & Lehmann, n.d; Quarto et al., 2014) and could potentially drive brain activity. The average level of brightness associated with each emotional stimuli (see Table 1), although it is not always significantly different, seems to follow the same pattern as the one observed for brain activity in the AC (Fear > Happy > sad/neutral). Experimentally manipulating brightness in future studies could help determine if this sound feature is crucial for emotion differentiation in the AC.

As for the similar activation observed in the amygdala for vocal and musical fear, the amygdala seem to have played its role of "relevance detector" for biologically significant events (Frühholz et al., 2016; Sander, Brechmann, & Scheich, 2003) by responding more to the fear stimuli. It is possible that the response to the amygdala for fear-expressing stimuli and not for those expressing other emotions can be explained by the short duration of stimuli and therefore the nature of the emotional information transmitted. Other studies have shown an increase in cerebral activity in the amygdala in the presentation of longer musical extracts expressing sadness (e.g. Mitterschiffthaler et al., 2007) or laughter and cries (Sander et al., 2003).

Taken together, regardless of the underlying causes, the observed emotion-specific activity irrespective of timbre for fear in the limbic structures and for fear, happiness, and sadness in the auditory cortex supports the proposal of a common neural network for both vocal and musical emotion (Peretz, 2010).

Conclusion

Using highly comparable and validated musical and vocal stimuli allowed us to circumvent stimuli comparability issues linked to specific timbre (music and voice) brain activity (respectively in the STG and STS) in a previous study (Aubé et al., 2015). Now that we have observed that both types of stimuli create a similar pattern of activation in the brain, these similar (vocal and musical) emotional stimuli can be used to determine how musical anhedonia (Mas-Herrero, Zatorre, & Rodriguez-Fornells, 2014), auditory affective agnosia (Heilman, Scholes, & Watson, 1975), amusia (Gosselin et al., 2015) and cochlear implants (Lehmann & Paquette, 2015) affect emotion perception across domains and help shed further light on the biological origins of music.

Rererences

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A. F., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex*, *59*, 126–137. http://doi.org/10.1016/j.cortex.2014.07.013

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*(3), 399–407. http://doi.org/10.1093/scan/nsu067

Bach, D., Schächinger, H., & Neuhoff, J. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex*, *18*(1), 145–50.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. http://doi.org/10.3758/BRM.40.2.531

Bestelmeyer, P., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *The Journal of Neuroscience*, *34*(24), 8098–8105.

Bestelmeyer, P., Rouger, J., DeBruine, L., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, *117*(2), 217–223.

Bigand, E., Filipic, S., & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy*, *1060*(1), 429–437.

Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*(3), 335–348. http://doi.org/10.1037/a0017928

Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. http://doi.org/10.1016/j.neuron.2007.10.004

Dellacherie, D., Hasboun, D., Baulac, M., Belin, P., & Samson, S. (2011). Impaired recognition of fear in voices and reduced anxiety after unilateral temporal lobe resection. *Neuropsychologia*, *49*(4), 618–629.

Escoffier, N., Zhong, J., & Schirmer, A. (2013). Emotional expressions in voice and music: same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations, *36*(2), 480–487. http://doi.org/10.1016/j.neuroimage.2007.02.043

Frühholz, S., Trost, W., & Kotz, S. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96–110.

Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex*, *71*, 171–182. http://doi.org/10.1016/j.cortex.2015.06.022

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244. http://doi.org/10.1016/j.neuropsychologia.2006.07.012

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., & Samson, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain : A Journal of Neurology*, *128*(Pt 3), 628–40. http://doi.org/10.1093/brain/awh420

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2141–2155.

Heilman, K. M., Scholes, R., & Watson, R. T. (1975). Auditory affective agnosia. Disturbed comprehension of affective speech. *Journal of Neurology, Neurosurgery, and Psychiatry*, *38*(1), 69–72. http://doi.org/10.1136/JNNP.38.1.69

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception*, *23*(4), 319–330. http://doi.org/10.1525/mp.2006.23.4.319

Juslin, P., & Laukka, P. (2003). Emotional expression in speech and music. *Annals of the New York Academy of*, *1000*(1), 279–282.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1812. http://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N. P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. http://doi.org/10.1037/0033-2909.129.5.770

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy, *81*, 49–60. http://doi.org/10.1016/j.neuroimage.2013.05.008

Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*, 1–8.

Laukka, P., & Juslin, P. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, *31*(3), 182–191.

Lehmann, A., & Paquette, S. (2015). Cross-domain processing of musical and vocal emotions in cochlear implant users. *Frontiers in Neuroscience*, *9*, 343. http://doi.org/10.3389/fnins.2015.00343

Mas-Herrero, E., Zatorre, R., & Rodriguez-Fornells, A. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology*, *24*(6), 699–704.

Mitterschiffthaler, M. M. T., Fu, C. H. Y. C., Dalton, J. A. J., Andrew, C. M., & Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, *28*(11), 1150–1162. http://doi.org/10.1002/hbm.20337

Paquette, S., Ahmed, D., Peretz, I., & Lehmann, A. (n.d.). Recognition of musical and vocal emotions through cochlear implants simulation. *Hearing Research*.

Paquette, S., Peretz, I., & Belin, P. (2013). The Musical Emotional Bursts: a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology*, *4*, 509. http://doi.org/10.3389/fpsyg.2013.00509

Peretz, I. (2010). Towards a Neurobiology of Musical Emotions. In J. Juslin, P; Sloboda (Ed.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 99–126). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199230143.003.0005

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. http://doi.org/10.1016/S0010-0277(98)00043-2

Quarto, T., Blasi, G., Pallesen, K. J., Bertolino, A., & Brattico, E. (2014). Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits. *PloS One*, *9*(7), e103278. http://doi.org/10.1371/journal.pone.0103278

Quarto, T., Blasi, G., Pallesen, K. J., Bertolino, A., Brattico, E., Fusar-Poli, P., … Lavie, N. (2014). Implicit Processing of Visual Emotions Is Affected by Sound-Induced Affective States and Individual Affective Traits. *PLoS ONE*, *9*(7), e103278. http://doi.org/10.1371/journal.pone.0103278

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(10), 3843–52. http://doi.org/10.1523/JNEUROSCI.4515-10.2011

Salimpoor, V. V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature*, *14*(2), 257–262. http://doi.org/10.1038/nn.2726

Sander, K., Brechmann, A., & Scheich, H. (2003). Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Research Protocols*, *11*(2), 81–91. http://doi.org/10.1016/S1385-299X(03)00018-7

Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *NeuroImage*, *63*(1), 137–147. http://doi.org/10.1016/j.neuroimage.2012.06.025

Scott, S., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Auditory recognition of emotion after amygdalotomy: impairment of fear and anger. *Nature*, *385*, 254–257.

Sharda, M., Midha, R., Malik, S., Mukerji, S., & Singh, N. C. (2015). Fronto-Temporal Connectivity is Preserved During Sung but Not Spoken Word Listening, Across the Autism Spectrum. *Autism Research*, *8*(2), 174–186. http://doi.org/10.1002/aur.1437

Sprengelmeyer, R., Young, A. W., Schroeder, U., Grossenbacher, P. G., Federlein, J., Buttner, T., & Przuntek, H. (1999). Knowing no fear, *266*(1437). http://doi.org/10.1098/rspb.1999.0945

Yik, M., Russell, J., & Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social*, *77*(3), 600.

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *11*(10), 946–953. http://doi.org/10.1093/cercor/11.10.946

Article 2: Tableaux et Figures

Tableau I. A2: Analyse acoustique des 120 stimuli (40 MAV, 80 MEB)

Acoustical analysis of the 120 stimuli (40 MAV, 80 MEB). Stimuli were analyzed for the presence of various acoustical features known to affect judgments (Juslin & Laukka, 2003; Quarto, Blasi, Pallesen, Bertolino, Brattico, et al., 2014).

Voices	Fear		P < 0.05	Happy		P < 0.05	Sad		P < 0.05	Neutral	
Brightness	0,45	(0,19)		0,40	(0,06)		0,37	(0,08)		0,36	(0,13)
RMS energy[1]	0,76	(0,15)	H, S	0,26	(0,07)	N	0,32	(0,15)	N	0,71	(0,16)
Roughness[n]	0,03	(0,04)		0,01	(0,01)	N	0,01	(0,00)	N	0,04	(0,02)
Min Pitch	150,89	(70,00)		148,46	(60,13)		235,93	(83,92)		153,65	(55,81)
Max Pitch	455,16	(156,19)		547,92	(78,96)		462,61	(110,46)	N	278,35	(177,93)
Mean Pitch	363,06	(81,37)	N	296,86	(97,20)	N	311,21	(83,15)	N	175,21	(70,62)
Temporal SIC[2]	0,22	(0,04)	H	0,32	(0,07)	N	0,26	(0,10)		0,21	(0,03)
Music	**Fear**			**Happy**			**Sad**			**Neutral**	
Brightness	0,64	(0,19)	S, N	0,50	(0,17)		0,40	(0,24)		0,44	(0,18)
RMS energy[1]	0,65	(0,27)	N	0,55	(0,17)	N	0,63	(0,13)	N	0,91	(0,20)
Roughness[n]	0,03	(0,05)		0,00	(0,00)		0,00	(0,00)		0,00	(0,01)
Min Pitch	279,34	(109,93)		261,47	(81,81)		281,41	(112,51)	N	351,25	(104,60)
Max Pitch	525,44	(133,88)	S, N	372,98	(140,77)	S, N	495,61	(130,60)		373,71	(115,35)
Mean Pitch	394,42	(105,61)		402,31	(116,12)		322,92	(97,40)		359,17	(106,54)
Temporal SIC[2]	0,25	(0,11)		0,21	(0,05)		0,20	(0,05)		0,21	(0,04)

Acoustical measure significantly different from (H = happy, S = Sad, N = Neural); Bonferroni corrected

[n] Roughness was normalised with minimum and maximum values of all stimuly.
[#] RMS energy (Xe^{-1}) and Rate of spectral information change (Xe^{-2}) values are presented using scienfic notation

Brightness (amount of energy above 3000 Hz; Juslin, 2000), energy (root mean square; RMS) and roughness (beating), as well as pitch values (min, mean, max) were extracted using the MIR toolbox (Lartillot & Toiviainen, 2007) and the rate of temporal cues (rate of Spectral Information Change (SIC): Rogalsky, Rong, Saberi, & Hickok, 2011 ; Sharda, Midha, Malik, Mukerji, & Singh, 2015) using MATLAB 2012Rb.

Figure 1. A2: Representation graphique des contrastes de deuxième niveau.

Second level group analysis contrasting emotional activity by timbre. Analysis revealed a main effect of emotion in the AC (blue: a, b) and in the AM/PHG (red: e,f) as well as a Timbre x Emotion interaction in the STG (green: c,d).

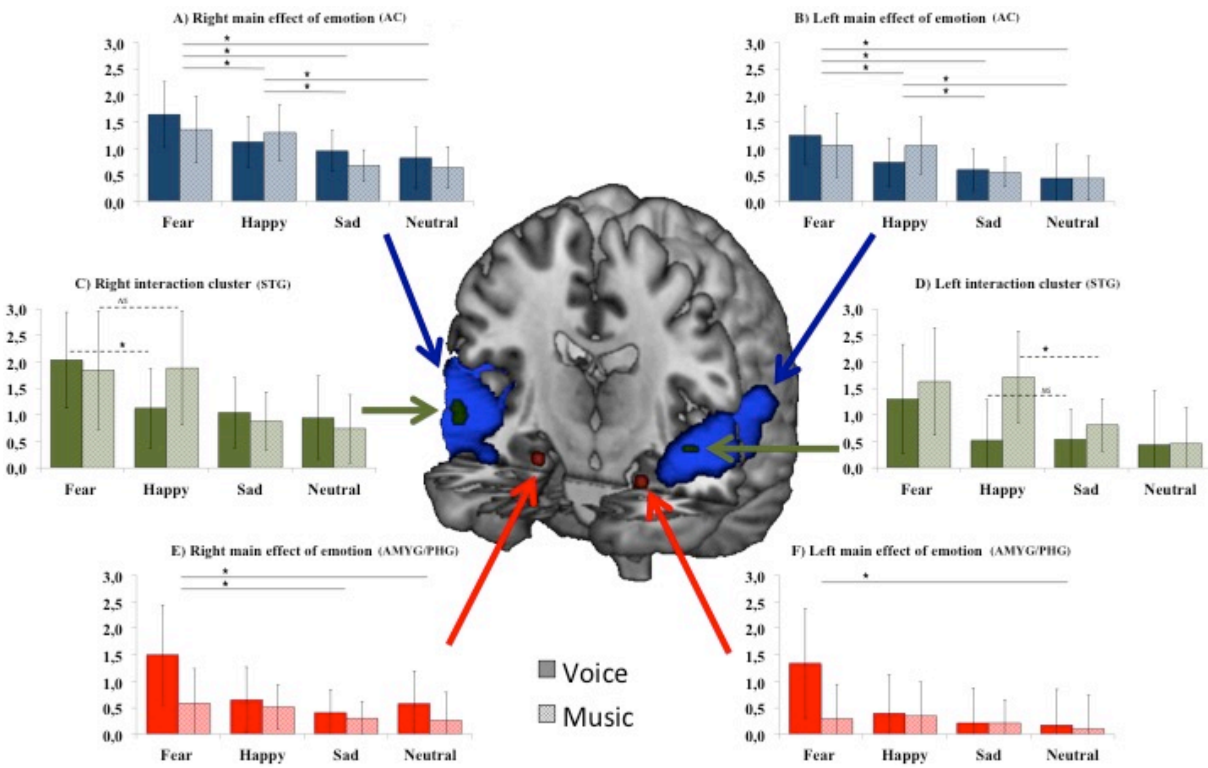
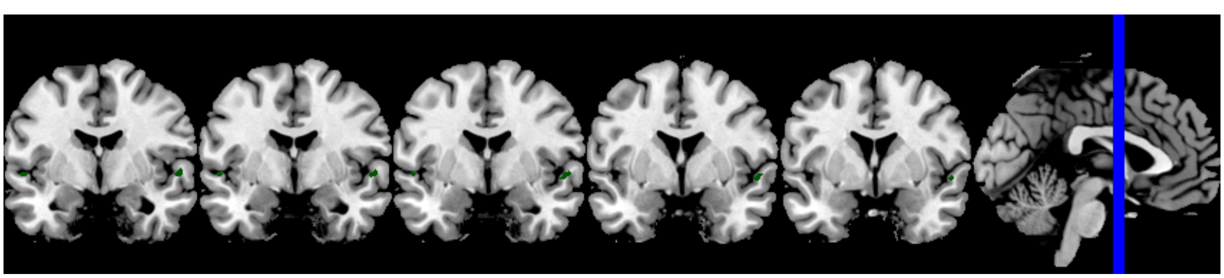


Figure 2. A2: Plan coronal représentant l’intéraction

Coronal representation of the Timbre (2) X Emotion (4) interaction [peak: -51 -4 -2; 51 -1 -5] in green, with represented slice position in blue.



A3: Recognition of musical and vocal emotions through cochlear implant simulation

Soumis à Hearing Research (2016)

Sébastien Paquette[1], Duha Ahmed[2], Isabelle peretz[1], Alexandre Lehmann[1,2]

[1]Department of Psychology, International Laboratory for Brain Music and Sound Research, Center for Research on Brain, Language and Music, University of Montreal, Montreal, QC, Canada

[2]Department of Otolaryngology Head and Neck Surgery, McGill University, Montreal, Québec, Canada

Keywords: Cochlear implants; emotional acoustical cues; cross-domain comparison; music; voice; timbre.

Musical and vocal emotions through cochlear implants simulation

Abstract

Cochlear implants can successfully restore hearing in profoundly deaf individuals and enable speech comprehension. The acoustic signal provided is severely degraded and, as a result, many important acoustic cues for perceiving emotion in voices and music are unavailable. The deficit of cochlear implant users in auditory emotion processing has been clearly established and has negative consequences for development and socio-professional integration. However, the exact extent of this deficit and the specific cues that remain available to cochlear implant users are unknown due to several confounding factors. Here we assessed the recognition of the most basic form of auditory emotion and aimed to identify which acoustic cues are most relevant to recognize emotions through cochlear implants. To do so, we used specific stimuli that allowed an assessment of basic auditory emotions (vocal, musical) while controlling for confounding factors. These stimuli, together with a validated cochlear implant simulation approach, allowed testing of natural versus cochlear implant hearing in the same participants. Results show that timbral acoustical cues (brightness, energy, and roughness) correlate with participant emotional rating for both vocal and musical emotion bursts in the cochlear implant simulation condition, suggesting that specific attention should be given to these cues in the design of cochlear implant processors and rehabilitation protocols, in order to improve emotion perception in this population. For instance, musical therapy focused on timbre could improve emotion perception and regulation, and thus improve social functioning in cochlear implant children during development.

Introduction

Cochlear implants (CI) can successfully restore some sense of hearing in profoundly deaf individuals. After intensive rehabilitation, most users can reach a good level of speech comprehension. However, the acoustic signal provided by the device is severely degraded, and the frequency resolution is poor. As a result, CI users are less accurate at discriminating pitch patterns (Gfeller & Lansing, 1992; Gfeller & Lansing, 1991; Hopyan et al., 2012) and identifying pitch changes or direction (Gfeller, Turner, et al., 2002; Laneau et al., 2004) than normal hearing controls.

Because understanding both vocal and musical emotional expressions require in part the processing of specific acoustic cues (Juslin & Laukka, 2003), most of which are based on pitch, CI users are impaired at perceiving those expressions. As opposed to normal hearing individuals who seem to rely on mean pitch (or mode: Hopyan, Manno III, Papsin, & Gordon, 2016), CI users are biased toward using pitch range cues (less salient cues: Gilbers et al., 2015). Perceiving emotion conveyed through vocal intonations (prosody) and through music is therefore highly challenging for CI users (Nakata et al., 2012; Wang et al., 2013). Fortunately, other non-pitch based cues can also convey emotions (Gabrielsson & Lindström, 2001; Gosselin et al., 2015) and these cues, such as temporal (e.g. rate, rhythm) and spectro-temporal (e.g. timbre) variations, are available to some extent to CI users (Gfeller, Witt, Mehr, et al., 2002; Kong et al., 2004; Looi et al., 2012) and reported to be used for emotion discrimination (e.g. tempo: Hopyan et al., 2016). CI users can discriminate happy from sad auditory emotions or identify them above chance, but not as well as normal hearing controls (Hopyan et al., 2012; Stabej, Smid, Gros, Zargi, Kosir, & Vatovec, 2012; Volkova et al., 2013). In a recent study by Ambert-Dahan and colleagues (2015), CI users identified musical emotions above chance level (happiness, sadness, threat and peacefulness), but demonstrated deficits in perceiving arousal (relaxing/stimulating) of musical excerpts, whereas ratings of valence (negative/ positive) remained unaffected. Judgments of emotional dimensions (arousal, valence) of stimuli do not seem to be uniformly impaired in CI users

Accurately perceiving emotions (or emotional dimensions) in language is crucial for social integration, interpersonal communication and a general quality of life. Access to musical emotions is also paramount: the most common reason for listening to music is its rich emotional content (Juslin & Sloboda, 2001; Lonsdale & North, 2011), and many deaf individuals specify music enjoyment as a major motivation for getting an implant (Gfeller, Turner, et al., 2002). This is all the more critical for developing children receiving an implant, as a failure to perceive their parents' or teacher's emotional expressions will result in maladjusted behaviour and the incapacity to express emotions in their own voice. This auditory emotional deficit can extend to general emotion processing. Although Hopyan-Misakyan and collaborators (2009) have shown that 10-year-old children who had received their CI prelingually were just as accurate as their normal hearing controls in discriminating between different facial expressions, normal hearing preschool children performed significantly better on facial expression recognition than prelingually implanted CI children (Wang, Su, Fang, & Zhou, 2011), suggesting that there is a delayed development in deaf children with CI. In contrast, Wiefferink and collaborators (2013; 2012) observed impaired emotional regulation (emotion expression and coping strategies) and social functioning (social competence and externalizing behaviours) in CI users compared to normal hearing individuals.

Whereas a deficit in auditory emotion processing has been clearly established in CI users, the extent of this deficit is unknown. A first issue is that several confounding factors may have caused an underestimation of the extent of the deficit because previous studies have used complex stimuli, either vocal (spoken sentences) or musical (songs). Factors such as semantics and language (voice), tempo (music), duration and context may involve the interaction of non-affective processing systems and contain redundancy that could make emotional recognition easier (Paquette et al., 2013). Hence, a necessary step to further study fundamental auditory emotion processing in CI users is to test basic emotions in their most primitive expressions (Lehmann & Paquette, 2015). Minimizing the role of confounding factors is also a key step in addressing two important issues in the literature. How CI users are able to perceive emotions (besides visual cues) despite their lack of access to many critical pitch-based cues has not been determined. Which cues do they rely on and what explains the large inter-individual variability? The second issue is concerned with the direct comparison of musical and vocal emotional processing in CI users. Music and voice are indeed hard to compare because of many confounding factors and also because voice has continuous pitch whereas most music studies on emotion employ the piano, which has a discrete pitch.

The goal of this study is to address these important questions by assessing the recognition of the most basic form of auditory emotion and comparing it between a cochlear implant and normal-hearing conditions. We aimed to identify which acoustic cues are most relevant for CI users to recognize emotion: correlations were made between emotional recognition and acoustical features to see which is used to extract auditory emotions when pitch perception is degraded. Finally, we sought to directly compare emotional recognition through cochlear implant simulations between the domains of music and voice.

To do so, we used validated sound stimuli that allow an assessment of the basic auditory emotions while controlling for confounding factors and allowing the most direct comparison between voice and music: the Montreal Affective Voices (MAV) and the Musical Emotional Bursts (MEB). The MAV (Belin et al., 2008) are non-speech vocalizations by ten professional actors (e.g., screams, laughter) depicting basic emotions that are fundamental to human communication (Scherer, 1986). They consist of short vocal interjections on the vowel /a/ expressing basic emotions (every actor expressed every emotion); they have minimal semantic information and minimal interaction with linguistic processes (Bestelmeyer et al., 2010). The MEB (Paquette et al., 2013) are designed to be the musical counterparts of the MAV and they consist of a few notes produced by professional musicians expressing basic musical emotions. They are all the more similar to vocal stimuli because they use continuous pitch instruments (clarinet or violin, which have a seamless progression between notes).

Because those stimuli have been previously validated, we can perform a well-controlled acoustical analysis and correlate the acoustical features with participants' emotional recognition. Acoustical cues (both pitch- and non-pitch based) known to affect emotional recognition in music and voice were quantified (computed) for each stimulus (Juslin & Laukka, 2003; Quarto et al., 2014): timbral cues (brightness, energy and roughness; Lartillot et al., 2007), pitch cues (minimum, maximum and average pitch; Boersma, 2002), and temporal cues (rate of Spectral Information Change (SIC), inspired by Rogalsky, Rong, Saberi, & Hickok, 2011; Sharda, Midha, Malik, Mukerji, & Singh, 2015: mean peak temporal modulation rate).

We use a validated cochlear implant simulation approach, allowing for testing of natural versus cochlear implant hearing in the same participants. Vocoding algorithms have been successfully employed to simulate the auditory signal degradation through a CI in normal hearing individuals; their use for perception research has been validated for both speech (Fu & Nogaki, 2005; Nogaki, Fu, & Galvin, 2007; Poissant, Whitmal, & Freyman, 2006; Qin & Oxenham, 2003) and music (Cousineau, Demany, Meyer, & Pressnitzer, 2010; Limb, 2006; Moore & Tan, 2003). Furthermore, this approach provides easier access to a larger number of homogenous participants (Driscoll, Oleson, Jiang, & Gfeller, 2009; Poissant et al., 2006). Because participants hear through cochlear implants for the first time, this has the added benefit of limiting the variability commonly observed in CI users, which is partly related to training outcomes.

Based on previous studies, we expect reduced emotion recognition accuracy and impaired arousal judgments in the cochlear implant simulation condition, when compared to the normal hearing condition. In terms of relevant acoustical features, we expect that the quantified value of certain acoustical cues that do not entirely rely on pitch (temporal cues and timbral cues) will correlate with participant judgments values.

Assessing the extent of the deficit using basic emotions and identifying available acoustical cues enabling CI users to optimally identify emotions can inform the optimization of implant processors and rehabilitation strategies to facilitate the development of emotion regulation strategies and social competence in this population. If successful, the next logical step shall be to apply this approach to a CI user population.

Method

Participants

Sixteen French and or English-speaking participants (10 females) between the ages of 23 and 52 (M=29) with no self-reported hearing problems were recruited through our laboratory participant mailing list and gave written informed consent. They had on average 18.5 years of education (SD 3.4) and 5.1 years of musical training (SD: 5.4). The ethics committee of the faculty of Arts and Sciences of the University of Montréal approved the following research protocol.

Stimuli

120 stimuli were used: 40 short (M=1.3 seconds) vocal emotional interjections from the MAV (Belin et al., 2008) and 80 (40 clarinet, 40 violin) short musical bursts (a few notes, M=1.6s) from the MEB (Paquette et al., 2013). Each category (voice, clarinet, and violin) contained ten stimuli per emotion (fear, sadness, happiness, neutrality). These 120 stimuli (original condition) were also presented in a CI-simulated version (CI-sim condition), processed through an 8-band noise vocoder developed to simulate CI-hearing in normal-hearing individuals. We used the cochlear implant simulation algorithm developed by Cousineau and collaborators (2010). This specific algorithm was selected because it equates the pitch discrimination capabilities of normal hearing individuals with those of CI users.

Procedure

Participants sat in a soundproof booth for a familiarization phase followed by the experimental protocol.

To familiarize participants with the hearing of CI-sim stimuli, they listened to vocal and music excerpts with and without CI-simulation. Participants were first presented with 11 spoken sentences from the IEEE database (Rothauser et al., 1969) if they spoke English (n=12), and from HINT database (Nilsson, Soli, & Sullivan, 1994) if they spoke French (n=4). After hearing the sentences in their CI-sim version, they were asked to repeat what they had heard and were then presented with the sentence in the original condition. Participants then heard seven CI-sim one-minute-long excerpts of popular music (e.g. Michael Jackson – Thriller, 1982). They were asked how familiar the song was on a scale from "not familiar at all" to "extremely familiar", and then asked if they could give us the title of the song. They then heard the original version.

For the main task, participants first heard a pseudo-random presentation of the 120 stimuli in their CI-sim version and then in their original version, in order to avoid an effect of prior exposition to the original version of the stimuli. They were asked, after hearing each excerpt, to rate them on seven different visual analog scales using a computer mouse. On the first page displaying four scales, they rated from "Absent" to "Present" related to how much a stimulus expressed each emotion (happiness, sadness, fear and/or neutrality). They then rated how confident they were about their emotional ratings from "Not at all confident" to "Extremely confident". On the second page, they rated the stimulus's emotional valence from "Extremely negative" to "Extremely positive", and its level of arousal from "Not at all arousing" to "Extremely arousing". This procedure has been adapted from Gosselin and colleagues (2007). The experiment was self-paced and the complete session lasted on average 90 minutes. During a practice run just before the main task, participants heard six example stimuli to ensure they understood the different rating dimensions and were familiarized with the rating interface. To minimize fatigue, participants were given short breaks for every 30 stimuli. The main task was programmed using the Psychtoolbox (Kleiner et al., 2007), the sound was delivered through BeyerDynamic DT 990 Pro headphones at 70 dB and rating scales were presented on an LG Flatron screen.

Analysis

Ratings on the four emotions scales (happy, sad, fear, neutral) were converted to accuracy scores, with the highest rating corresponding to the identified emotion (0.8% of all rating were excluded because two emotion were rated equally). To directly quantify the effect of the CI simulation on emotional ratings, difference scores were computed (original minus CI-simulated) for each participant. For accuracy scores, this subtraction was done between the average accuracy for each timbre and emotional categories. For arousal, valence and confidence ratings, this was done for each stimulus. Because the scores obtained for the confidence ratings showed different usage for each participant, these ratings were normalized using a Z-Score computed within participants to allow meaningful comparison. When applicable, a Holm–Bonferroni correction for multiple comparisons was applied.

The 240 stimuli (120 original, 120 CI-sim) were further analyzed and quantified for the presence of various acoustical features known to affect judgments (Juslin & Laukka, 2003; Quarto et al., 2014). Values for timber, defined by brightness (amount of energy above 3000 Hz; Juslin, 2000), energy (root mean square; RMS) and roughness (beating[4]), as well as pitch values (min, mean, max) and rate of temporal spectral information change were extracted to

[4] Interference pattern between two sounds of different frequencies, perceived as a periodic variation in volume, that has been believed to be more pronounced in dissonant than consonant sounds (Cousineau, McDermott, & Peretz, 2012).

allow us to identify the acoustical cues that were linked to the different emotional ratings (average values are presented in Table 1).

The acoustical analysis was performed on the stimuli using Matlab (Language & Computing, 2004), the MIRtoolbox (Lartillot & Toiviainen, 2007) and Praat (Boersma, 2002) Analyses were performed using SPSS (IBM Corp., Version 23.0. Armonk, NY) with a significance level of $p < .05$.

[Insert table 1 here]

Results

Accuracy

Stimuli were well recognized in their original condition (Table 2), except for the musical fear stimuli that were less accurately identified, similarly to what was observed in the validation study (Paquette et al., 2013). As expected, CI-sim stimuli were less well recognized than their original version. The CI simulation had a significant effect on all emotion identification; all differences (original minus CI-sim) were significantly ($p < 0.05$) different from 0 (null hypothesis). When presented through CI-simulation, vocal and musical fear stimuli were not recognized above chance (voice: $t(15) = 1.88$, $p > .05$; music: $t(15) = 1.92$, $p > .05$), neither were musical-sad stimuli ($t(15) = -.81$, $p > .05$; see Table 2).

In order to evaluate the effect of CI-simulation on emotion recognition accuracy, an ANOVA with Category (voice, music) and Emotion (happy, sad, fear, neutral) as between-subjects factors was computed with the difference in accuracy between the original and CI-sim ratings for each participant. A significant Category x Emotion interaction was observed (F(3,

45) = 13.88, p < .001, $\eta^2_{partial}$ = .48) as well as a main effect of Emotion (F(3, 45) = 8.75, p < .001, $\eta^2_{partial}$ = .37) but not of Category (F(1, 15) = 0.98, p = .34, $\eta^2_{partial}$ = .06).

Post hoc tests using the Bonferroni correction were used to break down the interaction and revealed which category of stimuli was most affected by the CI-simulation for each emotion. Musical happy stimuli were more affected by the CI-simulation than their vocal equivalent (p < .001) and the opposite was found for the fear stimuli as vocal stimuli were more affected their musical equivalent (p < .001; see Figure 1). Sad and neutral stimuli were not differently affected by the category of stimuli (all p > .05).

[Insert table 2 here] [Insert Figure 1 here]

Arousal and Valence Ratings

Valence and arousal ratings are less distinguishable when presented through CI-simulation; happy, sad and fear emotions are rated as more similar to neutral (Figure 2).

[Insert Figure 2 here]

Separate ANOVAs were computed for the arousal and valence ratings by considering the subtracted ratings (original minus CI-sim). For arousal, the ANOVA with Category (voice, music) and Emotion (happy, sad, fear, neutral) revealed a significant Category x Emotion interaction (F(3, 45) = 8.82, p < .001, $\eta^2_{partial}$ = .37) as well as a main effect of Emotion (F(3, 45) = 14.43, p < .001, $\eta^2_{partial}$ = .49), but not of Category (F(1, 15) = 0.33, p = .57, $\eta^2_{partial}$ = .02). By computing post hoc tests using the Bonferroni correction to compare the differences in arousal, we identified which category was more affected by the CI-simulation for each emotion. As can be seen in Figure 3, the arousal ratings for the music stimuli expressing happiness were significantly more affected by the CI-simulation than their vocal equivalent ($p < .01$). The other vocal and musical emotional stimuli were similarly affected (all $p > .05$).

The CI-simulation generally had no significant effect on the arousal ratings; only the differences (original minus CI-sim) for the fear stimuli (vocal: $t(15) = 4.76$, $p < .001$; musical $t(15) = 3.47$, $p < .05$) and musical-happy stimuli ($t(15) = 4.76$, $p < .001$) were significantly different from 0 (null hypothesis) wheras all others were not significantly affected (all $p > .5$).

[Insert Figure 3 here]

On valence ratings, the ANOVA yielded two main effects: Category: $F(1,15) = 101.78$, $p < .001$, $\eta^2_{partial} = .87$, and Emotion: $F(3, 45) = 53.63$, $p < .001$, $\eta^2_{partial} = .78$), which were significantly modulated by an interaction ($F(3, 45) = 13.26$, $p < .001$, $\eta^2_{partial} = .47$). Post hoc tests using the Bonferroni correction revealed that ratings of valence for negative vocal stimuli were significantly more affected by the CI-simulation than their musical counterpart (fear: $p < .001$; sad: $p < .001$) and the opposite was found for the happy stimuli as musical stimuli were more affected than their vocal equivalent ($p < .05$; see figure 4). Neutral vocal and musical stimuli were similarly affected (all $p = .75$).

The CI-simulation only significantly affected some of the valence ratings. Only the negative vocal stimuli (sad: $t(15) = -11.04$, $p < .001$; fear: $t(15) = -7.30$, $p < .001$) and the happy ($t(15) = 7.04$, $p < .001$) and neutral ($t(15) = 3.75$, $p < .05$) music stimuli were significantly different from 0 (null hypothesis), whereas all other differences (original minus CI-sim) in valence were not different from 0 (all $p > .07$).

[Insert Figure 4 here]

Confidence judgments

When performing a similar analysis to the ones above on participants' confidence judgment of their emotional ratings, only a main effect of Category (voice, music) was observed ($F(1,15) = 22.00$, $p < .001$, $\eta^2_{partial} = .60$), indicating that participants were generally less confident in their emotional ratings for the musical stimuli. This said, all confidence ratings for every emotion from every stimuli category were significantly affected by the procedure (all $p < 0.005$) and different from 0 (null hypothesis).

Correlation of emotional judgments with acoustical features

In order to determine which acoustical cues participants used to make their emotional judgment during the CI-sim condition, average ratings (accuracy, arousal, and valence) of each stimulus were correlated with each stimulus's previously extracted acoustic features: brightness, energy, roughness, pitch (min, mean, max) and rate of temporal spectral information (average values are presented in Table 1). Correlations by item (stimulus) were done separately for each category (voice, music) and version (original (O), CI-sim (CI)), but not by emotions in order to utilize the variety of valence and arousal values provided by different emotions.

Accuracy

Surprisingly, accuracy ratings correlated with no particular acoustical features of the stimuli, perhaps because emotion recognition is dependent on both arousal and valence appraisal; emotion recognition could be a result of a multidimensional analysis including both arousal and valence ratings (Bigand, Vieillard, et al., 2005) presented below.

Arousal

Brightness positively correlated with arousal ratings: the brighter a stimulus was (amount of energy above 3000 Hz), the more arousing the sound was rated, regardless of category and version (voice-O: $r(40) = .37$, $p < .05$; voice-CI: $r(40) = .34$, $p < .05$; music-O: $r(80) = .53$, $p < .001$; music-CI: $r(80) = .45$, $p < .001$).

Arousal ratings of the original musical stimuli were negatively correlated with their average amount of energy ($r(80) = -.41$, $p < .005$). This can be explained by the fact that the stimuli with the less variation in energy (neutral), had on average the highest energy (see Table 1). Similarly, when sounds (vocal or musical) were presented through the CI-simulation, stimuli energy (voice: $r(40) = -.46$, $p < .005$; music: $r(80) = -.33$, $p < .005$) negatively correlated with participants' arousal ratings.

Pitch cues mostly positively correlated with arousal judgments of the original vocal and musical stimuli, with maximum pitch (voice-O: $r(40) = .48$, $p < .005$; music-O: $r(80) = .62$, $p < .001$) and average pitch (voice-O: $r(40) = .67$, $p < .001$; music-O: $r(80) = .39$, $p < .001$). Interestingly, pitch values were also used to judge the CI-sim version; they positively correlated with the arousal judgements of the vocal (maximum pitch: $r(40) = .40$, $p < .05$; average pitch: $r(40) = .34$, $p < .05$;) stimuli and correlated negatively with the musical stimuli (minimum pitch: $r(80) = -.28$, $p < .05$).

The rate of temporal spectral information change only correlated positively with the arousal ratings of the musical stimuli in their original ($r(80) = .35$, $p < .001$) and CI-sim ($r(80) = .27$, $p < .05$) version, but surprisingly not the vocal stimuli (original: $r(40) = .27$, $p > .05$; CI-sim: $r(40) = .31$, $p > .05$).

Roughness (negatively correlated) was used to rate the specific version of each stimulus category. It was used to make arousal judgment of the original musical stimuli ($r(80) = .34$, $p < .005$) and the CI-sim vocal stimuli ($r(40) = -.40$, $p < .05$).

Valence

When sounds (vocal or musical) were presented through the CI-simulation, the stimuli's energy (voice: $r(40) = -.56$, $p < .001$; music: $r(80) = -.45$, $p < .001$) and roughness (voice: $r(40) = -.51$, $p < .005$; music: $r(80) = -.37$, $p < .001$)) negatively correlated with participant valence ratings.

As for the pitch cues, in this case, they were only used to rate the musical stimuli: maximum pitch ($r(80) = .31$, $p < .001$) and average pitch ($r(80) = .27$, $p < .05$) positively correlated with the valence ratings for the original musical stimuli and minimum pitch ($r(80) = -.23$, $p < .05$) for their CI-sim versions.

The rate of temporal spectral information change only positively correlated with valence ratings of the original vocal stimuli ($r(40) = .39$, $p < .05$).

In summary, pitch cues were used to make both arousal and valence judgments, but few were common to both voice and music in their CI-sim version. Furthermore, brightness and energy were related to arousal ratings, and roughness and energy to the valence ratings for musical and vocal stimuli in their CI-sim versions (for a summary of the correlations, see Table A.1 in Appendix A).

Discussion

We directly compared the perception of vocal and musical emotions through a cochlear implant simulation in healthy subjects using comparable short bursts of emotions. Our goal was to examine how emotion perception (recognition, valence, and arousal) is affected by cochlear implant listening to identify relevant cues accessible to CI users and directly compare how vocal and musical emotions are impacted. This careful examination also allowed us to identify that timbral (brightness, energy, and roughness) acoustical features are available to CI users to perceive both vocal and musical emotions.

All stimuli's emotion recognition accuracy was affected by the CI-simulation. Happy vocal stimuli were the most preserved. Musical fear stimuli were not severely affected by the CI-simulation but were generally poorly recognized.

As observed in the CI literature (e.g. Nakata et al., 2012), not all emotions were identified above chance under CI-simulation. In our case, fear and musical-sad CI-sim stimuli were not identified above chance by participants. Sad musical stimuli were often confused with another negative valence stimuli (fear) or perceived as neutral. Fear stimuli were perceived as neutral or confused with happy, as was sometimes the case for the vocal-sad stimuli. The confusion with happy (Volkova et al., 2013), as well as the high recognition accuracy for the happy stimuli (Hopyan et al., 2012; Stabej, Smid, Gros, Zargi, Kosir, & Vatovec, 2012), is similar to what has been observed in the CI user literature, thus validating our CI-simulation approach.

The stimuli used in our study allowed to measure variable degrees of valence (even if we only had one positive emotion) and arousal. All emotions could be distinguished along these dimensions. As reported by Paquette and collaborators (2013), for these stimuli, the arousal and valence ratings obtained here for the original stimuli fit well with this dimensional representation of emotions, with happy stimuli as conveying positive and arousing emotions, fear stimuli as conveying negative and arousing emotions (for the most part), sad stimuli as conveying moderately arousing and negative emotions, and the neutral stimuli as conveying an emotional valence that is neither positive nor negative with little arousal.

Perceived through CI simulation, all emotions were less differentiable; they tended to cluster more around the neutral position than in their original version. This neutralization effect was more pronounced for the valence ratings of the vocal-sad and music-neutral stimuli but seemed to generally affect the most arousing stimuli: music-happy, vocal fear stimuli (valence and arousal rating) and musical-fear stimuli (arousal ratings). This neutralization effect could explain the results of Ambert-Dahan and collaborators (2015), who reported that CI users had difficulty perceiving arousal in musical excerpts.

Some acoustical properties were used for both music and voice in their original form (brightness, mean-pitch, max-pitch, rate of temporal SIC). Again here the use of normal hearing individuals could account for the fact that participants still use pitch despite its degradation; our results cannot account for effects of deafness or adaptation to the device, such effects could be measured by applying our protocol to a population of CI users. Of interest here is that some acoustical properties of the stimuli were used for both categories (voice, music) in their CI-simulated form (timbral/spectro-temporal properties: brightness, energy, roughness) to make arousal and valence judgments. Timbral cues seem to be important when pitch perception is degraded, the same cues are known to be used by amusic individuals, who also experience a congenital pitch perception deficit, albeit in a milder form (e.g. Hyde & Peretz, 2004), to realize emotional judgments (Gosselin et al., 2015). Teaching CI users to focus on these acoustical properties could help emotion discrimination in the CI user population.

It is documented that CI users can to some extent differentiate instrumental timbres on qualitative scales (e.g. dull to brilliant; Gfeller, Witt, Mehr, et al., 2002). Of most relevance here is that CI users who undergo music rehabilitation demonstrate improved timbre identification from baseline with increases in their subjective appraisal of music (Gfeller, Witt, Adamek, et al., 2002). The current study proposes the idea that timbre discrimination (musical) training could also help emotion perception/discrimination in both the musical and vocal domains. Hence, musical therapy focused on timbre could help promote the improvement of optimal tools for emotional regulation and social functioning in the CI population during development.

The research strategy of using normal hearing individuals allowed us to compare the effect of CI processing within participants. However, it might have created a bias towards neutral. Although researchers have emphasized the benefit of cochlear implant stimulation (see Driscoll et al., 2009; Poissant et al., 2006), CI-simulated sounds can be perceived as quite neutral by normal-hearing individuals in comparison to their everyday perception of auditory emotions. Similarly, the fact that the participants were less confident about their judgment with the CI-sim stimuli could be attributed to the fact that distorted sound (musical and vocal) can be perceived by normal hearing individuals as less natural (Moore & Tan, 2003), such that their novelty could have lessened participant judgments. Perhaps reduced confidence in judgments to musical stimuli can be explained by the fact that we are, first and foremost, voice experts (Latinus & Belin, 2011; in addition to speech perception, we routinely extract from voices a wealth of socially-relevant information), and by our choice of stimuli, which had greater variability of expression/rendition for a specific emotion for music versus the same expression of an emotion for vocal sounds (e.g. laughter for joy).

Further neurobehavioural studies should be done with CI users, using a similar approach to compare musical and vocal emotions, first to assess how the mechanisms (neural markers) involved in vocal and musical emotional processing are affected by the implantation, and secondly, to directly evaluate the effects of timbre discrimination training on auditory emotion perception.

In sum, our results contribute to the characterization of the emotional perception impairment of cochlear implant users and bring forward common timbral acoustical cues (brightness energy and roughness) as instrumental for emotion perception (regardless of stimulus category) when pitch perception is degraded. This suggests that specific attention should be given to these cues in the design of cochlear implant processors and rehabilitation protocols in order to improve emotion perception in this population. For instance, musical therapy focused on timbre could improve emotion perception and regulation, thus bettering social functioning in CI children during development.

Acknowledgements

The project described was supported by a Research Incubator Award from the Centre for Research on Brain, Language and Music to AL and IP and by Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Awards from the Canadian Institute of Health Research to SP. We would like to thank Marion Cousineau, Olivier Dussault and Gabrielle Laufer for their help.

References

Ambert-Dahan, E., Giraud, A.-L., Sterkers, O., Samson, S., 2015. Judgment of musical emotions after cochlear implantation in adults with progressive deafness. Front. Psychol. 6, 181. doi:10.3389/fpsyg.2015.00181

Belin, P., Fillion-Bilodeau, S., Gosselin, F., 2008. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. Behav. Res. Methods 40, 531–539. doi:10.3758/BRM.40.2.531

Bestelmeyer, P., Rouger, J., DeBruine, L., Belin, P., 2010. Auditory adaptation in vocal affect perception. Cognition 117, 217–223.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., Dacquet, A., 2005. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. Cogn. 19, 1113–1139.

Boersma, P., 2002. Praat, a system for doing phonetics by computer. Glot Int. 5.9/10 5, 341–345. doi:10.1097/AUD.0b013e31821473f7

Cousineau, M., Demany, L., Meyer, B., Pressnitzer, D., 2010. What breaks a melody: Perceiving F0 and intensity sequences with a cochlear implant. Hear. Res. 269, 34–41. doi:10.1016/j.heares.2010.07.007

Driscoll, V.D., Oleson, J., Jiang, D., Gfeller, K., 2009. Effects of Training on Recognition of Musical Instruments Presented through Cochlear Implant Simulations. J. Am. Acad. Audiol. 20, 71–82.

Fu, Q.-J., Nogaki, G., 2005. Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing. J. Assoc. Res. Otolaryngol. 6, 19–27. doi:10.1007/s10162-004-5024-3

Gabrielsson, A., Lindström, E., 2001. The influence of musical structure on emotional expression. Oxford University Press.

Gfeller, K., Lansing, C., 1992. Musical Perception of Cochlear Implant Users as Measured by the Primary Measures of Music Audiation: An Item Analysis. J. Music Ther. 29, 18–39. doi:10.1093/jmt/29.1.18

Gfeller, K., Lansing, C.R., 1991. Melodic, Rhythmic, and Timbral Perception of Adult Cochlear Implant Users. J. Speech Lang. Hear. Res. 34, 916. doi:10.1044/jshr.3404.916

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J.F., Witt, S., Stordahl, J., 2002a. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. Cochlear Implants Int. 3, 29–53. doi:10.1002/cii.50

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., Ringgenberg, S., 2002b. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. J. Am. Acad. Audiol. 13, 132–45.

Gfeller, K., Witt, S., Mehr, M., Woodworth, G., 2002c. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. Ann. Otol. Rhinol. Laryngol. 111, 349–356.

Gilbers, S., Fuller, C., Gilbers, D., Broersma, M., Goudbeek, M., Free, R., Başkent, D., 2015. Normal-Hearing Listeners' and Cochlear Implant Users' Perception of Pitch Cues in Emotional Speech. Iperception. 6, 301006615599139. doi:10.1177/0301006615599139

Gosselin, N., Paquette, S., Peretz, I., 2015. Sensitivity to musical emotions in congenital amusia. Cortex 71, 171–182. doi:10.1016/j.cortex.2015.06.022

Gosselin, N., Peretz, I., Johnsen, E., Adolphs, R., 2007. Amygdala damage impairs emotion recognition from music. Neuropsychologia 45, 236–244. doi:10.1016/j.neuropsychologia.2006.07.012

Hopyan-Misakyan, T., Gordon, K., Dennis, M., 2009. Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. Child Neuropsychol. 15, 136–46.

Hopyan, T., Manno III, F.A., Papsin, B.C., Gordon, K.A., 2016. Sad and happy emotion discrimination in music by children with cochlear implants. Child Neuropsychol. 22, 366–380. doi:10.1080/09297049.2014.992400

Hopyan, T., Peretz, I., Chan, L.P., Papsin, B.C., Gordon, K.A., 2012. Children Using Cochlear Implants Capitalize on Acoustical Hearing for Music Perception. Front. Psychol. 3, 425. doi:10.3389/fpsyg.2012.00425

Hyde, K.L., Peretz, I., 2004. Brains That Are out of Tune but in Time. Psychol. Sci. 15, 356–360. doi:10.1111/j.0956-7976.2004.00683.x

Juslin, P.N., 2000. Cue utilization in communication of emotion in music performance: Relating performance to perception. J. Exp. Psychol. Hum. Percept. Perform. 26, 1797–1812. doi:10.1037/0096-1523.26.6.1797

Juslin, P.N. (Ed), Sloboda, J.A. (Ed), 2001. Music and emotion: Theory and research. Oxford University Press.

Juslin, P.N.P., Laukka, P., 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol. Bull. 129, 770–814. doi:10.1037/0033-2909.129.5.770

Kleiner, M., Brainard, D.H., Pelli, D.G., Broussard, C., Wolf, T., Niehorster, D., 2007. What's new in Psychtoolbox-3? Perception 36, S14. doi:10.1068/v070821

Kong, Y.-Y., Cruz, R., Jones, J.A., Zeng, F.-G., 2004. Music perception with temporal cues in acoustic and electric hearing. Ear Hear. 25, 173–85.

Laneau, J., Wouters, J., Moonen, M., 2004. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. J. Acoust. 3606–3619.

Language, T., Computing, T., 2004. MATLAB The Language of Technical Computing. Components 3, 750. doi:10.1007/s10766-008-0082-5

Lartillot, O., Toiviainen, P., 2007. A matlab toolbox for musical feature extraction from audio. Int. Conf. Digit. Audio Eff. 1–8.

Latinus, M., Belin, P., 2011. Human voice perception. Curr. Biol. 21, R143–R145. doi:10.1016/j.cub.2010.12.033

Lehmann, A., Paquette, S., 2015. Cross-domain processing of musical and vocal emotions in cochlear implant users. Front. Neurosci. 9, 343. doi:10.3389/fnins.2015.00343

Limb, C.J., 2006. Cochlear implant-mediated perception of music. Curr. Opin. Otolaryngol. Head Neck Surg. 14, 337–340. doi:10.1097/01.moo.0000244192.59184.bd

Lonsdale, A.J.A., North, A.C.A., 2011. Why do we listen to music? A uses and gratifications analysis. Br. J. Psychol. 102, 108–134. doi:10.1348/000712610X506831

Looi, V., Gfeller, K., Driscoll, V., 2012. Music Appreciation and Training for Cochlear Implant Recipients: A Review. Semin. Hear. 33, 307–334. doi:10.1055/s-0032-1329222

Moore, B.C.J., Tan, C.-T., 2003. Perceived naturalness of spectrally distorted speech and music. J. Acoust. Soc. Am. 114, 408. doi:10.1121/1.1577552

Most, T., Aviner, C., 2009. Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing AIDS, and normal hearing. J. Deaf Stud. Deaf Educ. 14, 449–64. doi:10.1093/deafed/enp007

Nakata, T., Trehub, S.E., Kanda, Y., 2012. Effect of cochlear implants on children's perception and production of speech prosody. J. Acoust. Soc. Am. 131, 1307. doi:10.1121/1.3672697

Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 95, 1085–99.

Nogaki, G., Fu, Q.-J., Galvin, J.J., 2007. Effect of training rate on recognition of spectrally shifted speech. Ear Hear. 28, 132–40. doi:10.1097/AUD.0b013e3180312669

Paquette, S., Peretz, I., Belin, P., 2013. The Musical Emotional Bursts: a validated set of musical affect bursts to investigate auditory affective processing. Front. Psychol. 4, 509. doi:10.3389/fpsyg.2013.00509

Poissant, S.F., Whitmal, N.A., Freyman, R.L., 2006. Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. J. Acoust. Soc. Am. 119, 1606. doi:10.1121/1.2168428

Qin, M.K., Oxenham, A.J., 2003. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J. Acoust. Soc. Am. 114, 446. doi:10.1121/1.1579009

Quarto, T., Blasi, G., Pallesen, K.J., Bertolino, A., Brattico, E., 2014. Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits. PLoS One 9, e103278. doi:10.1371/journal.pone.0103278

Rogalsky, C., Rong, F., Saberi, K., Hickok, G., 2011. Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. J. Neurosci. 31, 3843–52. doi:10.1523/JNEUROSCI.4515-10.2011

Rothauser, E.H., Chapman, W.D., Guttman, N., Hecker, M.H.L., Nordby, K.S., Silbiger, H.R., Urbanek, G.E., Weinstock, M., 1969. IEEE Recommnded Practice for Speech Quality Measurements. IEEE Trans. Audio Electroacoust. 17, 225–246. doi:10.1109/TAU.1969.1162058

Scherer, K.R., 1986. Vocal affect expression: a review and a model for future research. Psychol. Bull. 99, 143–165. doi:10.1037/0033-2909.99.2.143

Sharda, M., Midha, R., Malik, S., Mukerji, S., Singh, N.C., 2015. Fronto-Temporal Connectivity is Preserved During Sung but Not Spoken Word Listening, Across the Autism Spectrum. Autism Res. 8, 174–186. doi:10.1002/aur.1437

Stabej, K.K., Smid, L., Gros, A., Zargi, M., Kosir, A., Vatovec, J., 2012. The music perception abilities of prelingually deaf children with cochlear implants. Int. J. Pediatr. Otorhinolaryngol. 76, 1392–1400. doi:10.1016/j.ijporl.2012.07.004

Volkova, A., Trehub, S.E., Schellenberg, E.G., Papsin, B.C., Gordon, K.A., 2013. Children with bilateral cochlear implants identify emotion in speech and music. http://dx.doi.org/10.1179/1754762812Y.0000000004 14, 80–91.

Wang, D.J., Trehub, S.E., Volkova, A., van Lieshout, P., 2013. Child implant users' imitation of happy- and sad-sounding speech. Front. Psychol. 4. doi:10.3389/fpsyg.2013.00351

Wang, Y., Su, Y., Fang, P., Zhou, Q., 2011. Facial expression recognition: Can preschoolers with cochlear implants and hearing aids catch it? Res. Dev. Disabil. 32, 2583–2588.

Wiefferink, C.H., Rieffe, C., Ketelaar, L., De Raeve, L., Frijns, J.H.M., 2013. Emotion understanding in deaf children with a cochlear implant. J. Deaf Stud. Deaf Educ. 18, 175–86. doi:10.1093/deafed/ens042

Wiefferink, C.H., Rieffe, C., Ketelaar, L., Frijns, J.H.M., 2012. Predicting social functioning in children with a cochlear implant and in normal-hearing children: The role of emotion regulation. Int. J. Pediatr. Otorhinolaryngol. 76, 883–889. doi:10.1016/j.ijporl.2012.02.065

Xin Luo, X., Fu, Q.-J., Galvin, J.J., 2007. Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users. Trends Amplif. 11, 301–315. doi:10.1177/1084713807305301

Article 3: Tableaux et Figures

Tableau I. A3: Caractéristiques acoustiques des extraits présentés

Average value of extracted acoustical features presented for each stimuli type (voice, Music), emotion (happy, sad, fear neutral) and condition (original, CI-sim) standard deviation is presented in parentheses.

Original	Voice				Music			
Acoustical values	Happy	Sad	Fear	Neutral	Happy	Sad	Fear	Neutral
Brightness	0,40 (0,06)	0,37 (0,08)	0,45 (0,19)	0,36 (0,13)	0,50 (0,17)	0,40 (0,24)	0,64 (0,19)	0,44 (0,18)
RMS energy[1]	0,26 (0,07)	0,32 (0,15)	0,76 (0,15)	0,71 (0,16)	0,55 (0,17)	0,63 (0,13)	0,65 (0,27)	0,91 (0,20)
Roughness[n]	0,01 (0,01)	0,01 (0,00)	0,03 (0,04)	0,04 (0,02)	0,00 (0,00)	0,00 (0,00)	0,03 (0,05)	0,00 (0,01)
Min Pitch	150,89 (70,00)	148,46 (60,13)	235,93 (83,92)	153,65 (55,81)	279,34 (109,93)	261,47 (81,81)	281,41 (112,51)	351,25 (104,60)
Max Pitch	455,16 (156,19)	547,92 (78,96)	462,61 (110,46)	278,35 (177,93)	525,44 (133,88)	372,98 (140,77)	495,61 (130,60)	373,71 (115,35)
Mean Pitch	296,86 (97,20)	311,21 (83,15)	363,06 (81,37)	175,21 (70,62)	402,31 (116,12)	322,92 (97,40)	394,42 (105,61)	359,17 (106,54)
Rate of temporal SIC[2]	0,32 (0,07)	0,26 (0,10)	0,22 (0,04)	0,21 (0,03)	0,21 (0,05)	0,20 (0,05)	0,25 (0,11)	0,21 (0,04)
CI-simulated	Voice				Music			
Brightness	0,51 (0,05)	0,46 (0,05)	0,51 (0,19)	0,38 (0,09)	0,56 (0,16)	0,48 (0,21)	0,68 (0,16)	0,48 (0,20)
RMS energy[1]	0,26 (0,07)	0,32 (0,15)	0,76 (0,15)	0,71 (0,16)	0,55 (0,17)	0,63 (0,13)	0,65 (0,27)	0,91 (0,20)
Roughness[n]	0,03 (0,02)	0,06 (0,07)	0,34 (0,13)	0,28 (0,13)	0,18 (0,13)	0,20 (0,06)	0,33 (0,27)	0,48 (0,17)
Min Pitch	84,10 (8,77)	76,42 (6,08)	76,77 (2,92)	82,15 (21,60)	78,22 (6,70)	110,32 (113,67)	94,91 (58,97)	175,25 (159,30)
Max Pitch	550,12 (68,87)	576,93 (27,96)	323,18 (204,32)	325,41 (190,91)	481,08 (159,31)	533,28 (101,89)	463,95 (143,50)	438,74 (173,57)
Mean Pitch	252,48 (65,62)	266,44 (72,88)	164,11 (82,31)	164,84 (107,28)	244,50 (106,55)	259,82 (134,75)	235,14 (126,78)	266,69 (169,54)
Rate of temporal SIC[2]	0,31 (0,02)	0,27 (0,07)	0,30 (0,04)	0,26 (0,03)	0,32 (0,04)	0,30 (0,06)	0,31 (0,05)	0,31 (0,06)

[n] Roughness was normalised with minimum and maximum values of all stimuli.
[#] RMS energy (Xe^{-1}) and Rate of spectral information change (Xe^{-2}) values are presented using scienfic notation

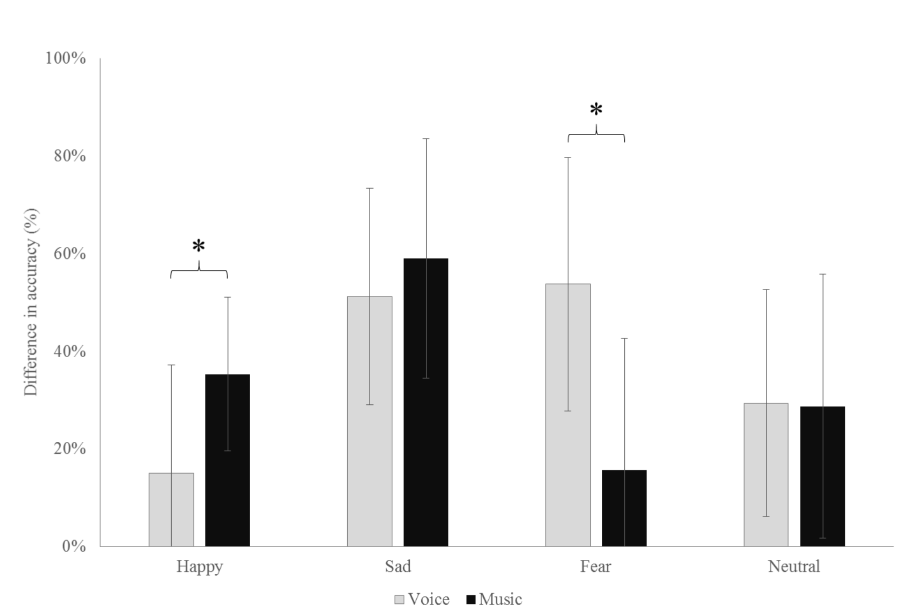
Tableau II. A3: Pourcentage d'identification (stimuli originaux et IC simulé)

Confusion matrix of emotion recognition (accuracy %) for the vocal and musical sounds in both the CI-sim and original conditions. When two emotions were rated equally, identification was rated undefined. Framed values were not significantly different from chance level (25%).

Accuracy: Original	Voice					Music				
Intended emotions	Happy	Sad	Fear	Neutral	*Undefined*	Happy	Sad	Fear	Neutral	*Undefined*
Happy	96.9	1.3	0.6	0.6	0.6	86.6	1.6	2.2	9.1	0.6
Sad	5.6	92.5	1.3	0.6	-	5.9	80.0	2.2	11.3	0.6
Fear	1.3	4.4	90	4.4	-	25.6	8.4	47.8	17.5	0.6
Neutral	3.1	1.9	1.9	93.1	-	6.3	18.4	1.6	73.4	0.3
Accuracy: ci-simulated	Voice					Music				
Happy	81.9	-	6.9	8.1	3.1	51.3	2.2	17.8	27.5	1.3
Sad	32.5	41.3	10.6	15.0	0.6	11.9	$\boxed{20.9}$	36.6	29.1	1.6
Fear	10.0	7.5	$\boxed{36.3}$	45.0	1.3	22.2	4.1	$\boxed{32.2}$	41.6	-
Neutral	6.3	11.3	16.9	63.8	1.9	4.4	13.1	36.6	44.7	1.3

Figure 1. A3: Effet de la simulation d'IC sur la reconnaissance d'émotions

Average effect (difference) of CI simulation on accuracy scores for each emotion (happy, sad, fear neutral) and category (voice, music). Error bars represent the standard deviation.



* = $p < .05$

Figure 2. A3: Jugements d'arousal et de la valence (stimuli originaux et IC simulé)

Raw arousal and valence ratings for both the vocal and musical stimuli presented for each emotion (happy, sad, fear neutral) and condition (original, CI-sim).

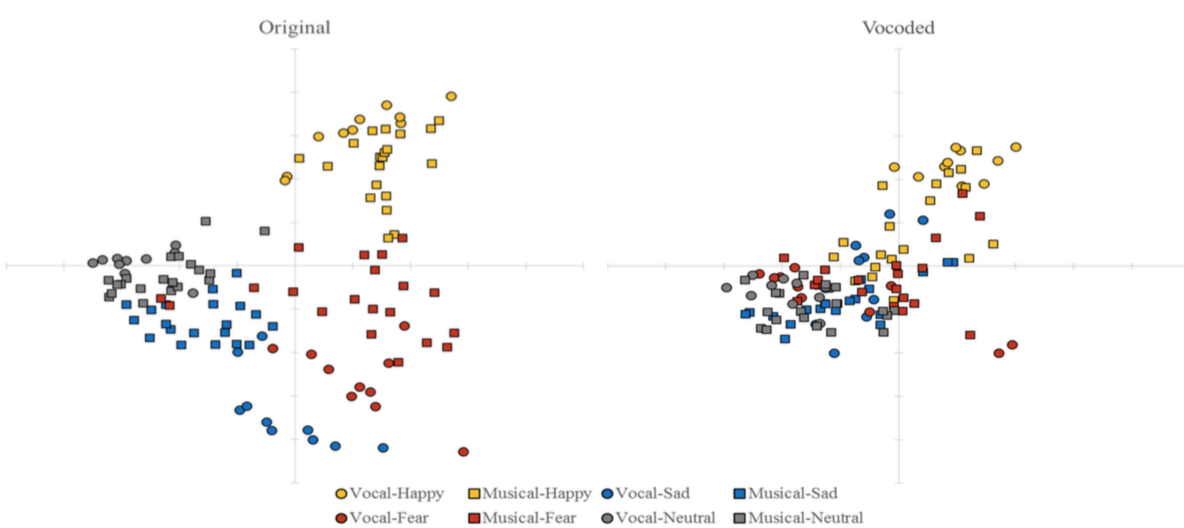
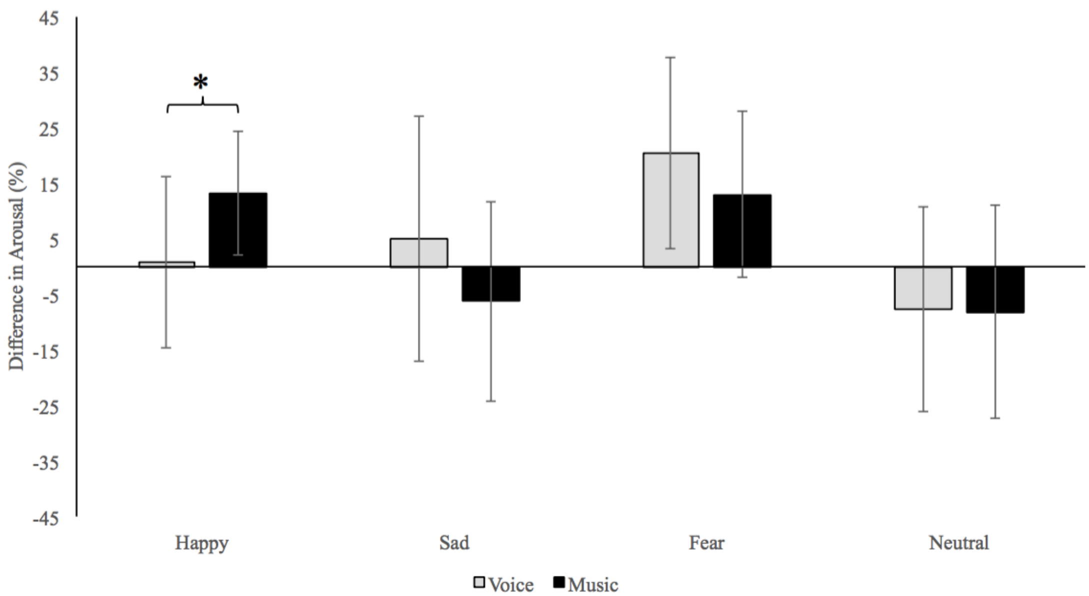


Figure 3. A3: Effet de la simulation d'IC sur les jugements d'arousal

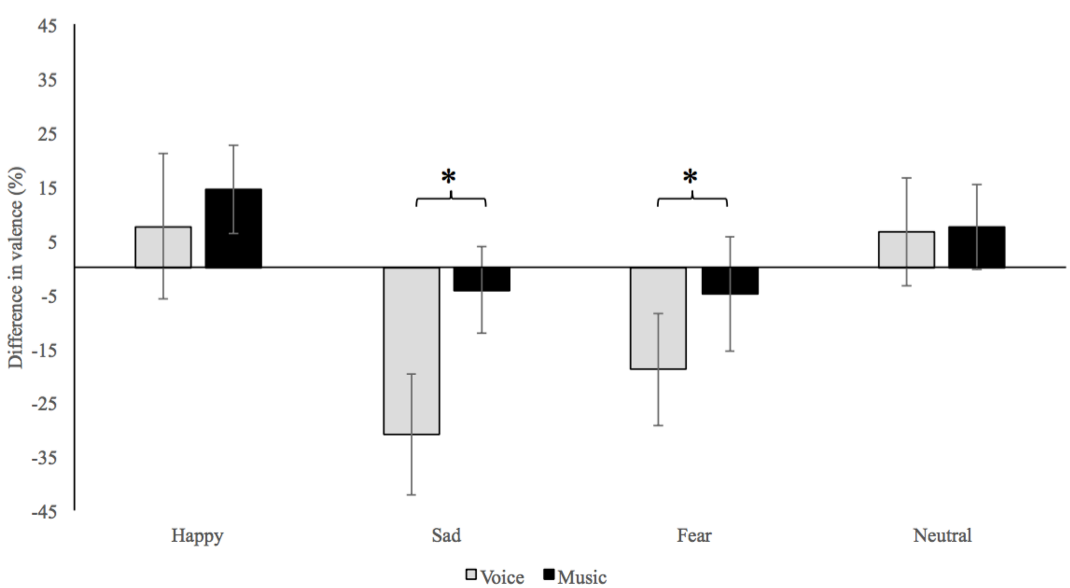
Average effect of the CI simulation on arousal ratings presented for each emotion (happy, sad, fear neutral) and stimuli type (voice, music). Error bars represent the standard deviation.



* = $p < .05$

Figure 4. A3: Effet de la simulation d'IC sur les jugements de la valence

Average effect of the CI simulation on valence ratings presented for each emotion (happy, sad, fear neutral) and stimuli type (voice, music). Error bars represents the standard deviation.



* = $p < .05$

Tableau III. A3: Matériel supp. 1: Corrélations avec les caractéristiques acoustiques

Pearson correlation values between extracted acoustical features and participants emotional ratings presented for each stimuli type (voice (40), Music (80)) and condition (original, CI-simulation). Shaded values are significant correlation common to vocal and musical stimuli in their CI-simulated version.

Original	Voice			Music		
Acoustical values	Accuracy	Valence	Arousal	Accuracy	Valence	Arousal
Brightness	0.12	-0.06	0.37*	-0.15	-0.03	0.53**
RMS energy	-0.18	-0.30	-0.26	-0.10	-0.09	-0.41**
Roughness	0.17	-0.11	-0.19	-0.60	-0.16	0.34**
Min Pitch	-0.19	-0.21	0.25	-0.12	0.08	-0.09
Max Pitch	-0.05	-0.24	0.48**	0.06	0.31**	0.62**
Mean Pitch	-0.02	-0.25	0.67**	-0.03	0.27*	0.39**
Rate of temporal SIC[2]	0.01	0.39*	0.27	0.03	-0.14	0.35**
CI-simulated	Voice			Music		
Brightness	0.01	0.18	.34*	-.07	0.22	0.45**
RMS energy	-0.21	-0.56**	-.46**	0.06	-0.45**	-0.33**
Roughness	-0.17	-0.51**	-0.40*	0.06	-0.37**	-0.14
Min Pitch	0.19	0.11	0.03	0.02	-0.23*	-0.28*
Max Pitch	0.11	0.28	0.40**	0.05	0.11	0.10
Mean Pitch	0.13	0.17	0.34*	0.01	-0.01	-0.06
Rate of temporal SIC[2]	0.01	0.28	0.31	-0.11	0.17	0.27*

** = p < .05

* = p < .01

Chapitre 3: Discussion Générale

3.1 Rappel des objectifs et intégration des résultats

L'objectif de cette thèse était d'apporter un support à la théorie que la musique recycle les circuits émotionnels qui ont évolué principalement pour le traitement des vocalisations biologiquement importantes (Peretz, 2010); que la musique s'est développée *à partir des rythmes et cadences de discours passionnés* (Spencer, 1857). Les prédispositions biologiques (reconnues dès un très jeune âge, de manière universelle et de façon distincte: Peretz, 2010) et les corrélats neuronaux similaires observés pour la perception des émotions vocales et musicales, combinés à leur grande similarité acoustique semblent tous supporter cette hypothèse. Il demeure que certaines manipulations acoustiques affectent différemment les émotions musicales et vocales (Ilie & Thompson, 2006) et que certains corrélats neuronaux semblent spécifiques aux stimuli musicaux et vocaux (Aubé et al., 2015; Escoffier et al., 2013). Il pourrait être stipulé que ces différences observées sont liées aux stimuli utilisés dans ces études. Dans cette thèse, de manière à minimiser ces différences observées, la comparaison des émotions musicales et vocales a été réalisée, en contrôlant pour les facteurs qui pourraient différencier ces deux types de stimuli (p.ex. longueur, complexité).

Pour atteindre l'objectif global de la thèse, les travaux qui la sous-tendent ont dû se réaliser en deux temps.

Dans un premier temps, l'objectif était de développer les outils pour mener notre projet à bien. Pour ce faire, il a fallu développer une batterie de stimuli émotionnels musicaux comparables aux stimuli vocaux validés déjà disponibles. Le but était de créer une batterie d'éclats émotionnels musicaux exprimant la joie, la peur, la tristesse et la neutralité, conçue pour être l'équivalent des Voix affectives Montréalaises (MAV; Belin et al., 2008). Seules trois émotions (et la neutralité) ont été sélectionnées pour permettre une comparaison directe entre les émotions vocales et musicales de base. Notre sélection réduite d'émotions limite la comparaison voix/musique à celles-ci, mais a permis une première comparaison contrôlée.

En dépit de leur courte durée, les émotions exprimées par les Éclats Émotionels Musicaux (MEB) ont été correctement identifiées et les stimuli ont obtenu des jugements d'arousal (éveil/stimulation) et de valence correspondant à l'émotion qu'il représentait. Ces stimuli interprétés au violon et à la clarinette peuvent donc être considérés comme une forme primitive d'émotions musicales, située quelque part entre des compositions musicales (p.ex. Aubé et al., 2015; Dalla Bella et al., 2003; Filipic et al., 2010; Peretz et al., 1998) et des tons modulés en fréquence de synthèse conçus pour imiter les caractéristiques acoustiques d'expressions vocales humaines (Kantrowitz et al., 2011). Les MEB représentent donc la forme la plus élémentaire de l'émotion musicale qui peut être étroitement liée aux expressions vocales, tout en restant naturels (réalistes; non synthétiques).

Dans un deuxième temps, nous avons utilisé les éclats musicaux et les MAV dans deux études expérimentales. Premièrement, pour effectuer une comparaison des circuits neuronaux utilisés pour le traitement des émotions musicales et vocales. Pour ce faire, une étude en imagerie par résonnance magnétique fonctionnelle a été réalisée. Deuxièmement, nous avons directement comparé par simulation la perception des émotions vocales et musicales dans une population (les implantés cochléaires : IC) chez qui certains indices acoustiques liés à la perception émotionnelle (p.ex. les hauteurs tonales) sont grandement affectés.

Dans ces études, les stimuli performés sur le violon et la clarinette ont été combinés sous l'étiquette musique. Premièrement, afin de faciliter la présentation des résultats et la comparaison avec la littérature. Deuxièmement, pour comparer les stimuli vocaux à un ensemble plus grand de stimuli musicaux, représentant de manière plus réelle la diversité d'émotions musicales auxquelles nous sommes exposés; l'instrument sur lequel la musique est jouée est connu pour avoir un impact sur la reconnaissance des émotions (Balkwill & Thompson, 1999; Behrens & Green, 1993; Gabrielsson & Juslin, 1996; Hailstone et al., 2009). Cette combinaison des éclats musicaux à la clarinette et au violon sous l'étiquette musique nous a également permis d'accentuer le parallèle avec la banque de stimuli vocaux qui contient des vocalisations de femmes et d'hommes à parts égales. La discrimination du genre dans la voix comme la discrimination d'instruments musicaux repose majoritairement sur le timbre (Pernet & Belin, 2012).

Lors de l'étude en Imagerie par Raisonnance Mangnétique fonctionelle, les zones d'activations cérébrales observées pour les émotions vocales et musicales ont montré des similitudes frappantes, indiquées par l'absence d'effet principal des timbres (Voix, Musique). De manière générale, les émotions auditives (joie, peur, tristesse) se sont révélées être associées à des niveaux d'activité cérébrale différenciables dans le cortex auditif, indépendamment de leur timbre. Les stimuli de peur musicaux et vocaux ont également suscité beaucoup plus d'activation que les autres émotions dans les structures limbiques inférieures (p.ex. amygdale).

Une exception a été observée dans une région circonscrite du Gyrus Temporal Supérieur (GTS) bilatéralement ou les stimuli musicaux exprimant la joie ont été associés à une plus grande activité cérébrale (similaire à celle des stimuli de peur) que les stimuli vocaux exprimant la joie. Cette différence d'activité cérébrale observée entre les stimuli musicaux et vocaux de joie pourrait, malgré les efforts déployés pour la validation de stimuli, être associée à une différence de valence émotionnelle. Lors de la validation, les stimuli musicaux et vocaux exprimant la joie ont tous deux été jugés comme étant positif et stimulant (certains stimuli musicaux ont été jugés comme étant les plus stimulants), mais les stimuli musicaux ont été jugés moins positifs que leurs équivalents vocaux (des rires). À tel point que les stimuli de joie exprimés sur le violon (lorsqu'ils étaient mal identifiés) étaient confondus avec des stimuli de peur (stimulant, mais négatif). La même confusion pourrait avoir eu lieu dans cette région circonscrite du GTS ou les stimuli de joie musicaux ont été associés à une activité similaire aux stimuli de peur. Cette région pourrait être associée plus spécifiquement au traitement de l'arousal (traiter préférentiellement les sons stimulants) en ne tenant pas compte de leur valence. Elle pourrait réaliser un traitement similaire à celui impliqué dans la régulation du système nerveux sympathique (ou recevoir cette information de l'insula), il a été démontré que les mesures de conductance électrodermale enregistrées lors de présentation de stimuli musicaux varient systématiquement avec le niveau d'arousal perçu dans le stimulus (Khalfa, Peretz, Blondin, & Robert, 2002). Dans deux études antérieures (Aubé et al., 2015; Escoffier et al., 2013) comparant l'activité cérébrale liée aux stimuli émotionnels vocaux et musicaux, cette région antérieure du GTS a été identifiée comme étant particulièrement sensible à la musique (en comparaison aux expressions vocales). Il est possible que cette plus grande sensibilité observée pour les stimuli musicaux soit due au fait que les stimuli musicaux utilisés dans ces études étaient plus stimulants que les vocaux (comme il est observé pour la joie dans l'article 2 de cette thèse). Leurs stimuli musicaux auraient agi comme une voix super expressive (Juslin & Laukka, 2003; Juslin & Västfjäll, 2008); les stimuli musicaux pourraient avoir agi comme une caricature de la voix en exagérant certaines caractéristiques des expressions vocales non linguistiques. Dans notre étude, mis à part cette exception liée aux stimuli musicaux de joie, les différents niveaux d'activation pour chaque émotion (peur, joie, tristesse; indifféremment du timbre) dans le cortex auditif sont conformes à la proposition d'une voie neuronale commune pour le traitement des émotions vocales et musicales.

Cette hypothèse est également supportée par l'effet principal d'émotion observé pour les stimuli de peur dans les structures limbiques inférieures. Pour la voix et la musique, l'amygdale viendrait ici jouer son rôle de «détecteur de pertinence» pour les évènements biologiquement significatifs (Frühholz et al., 2016; Sander et al., 2003) en répondant pour les stimuli de peur: musicaux et vocaux. Il est possible que la réponse à l'amygdale pour les stimuli exprimant la peur et non pour ceux exprimant les autres émotions soit explicable par la courte durée des stimuli et par conséquent la nature de l'information émotionnelle transmise. D'autres études ont en effet montré une augmentation de l'activité cérébrale dans l'amygdale lors de la présentation d'extraits musicaux plus longs évoquant la tristesse (p.ex. Mitterschiffthaler et al., 2007). Le cortex auditif viendrait ici décoder tel que décrit par Frühholz et collaborateurs (2016) les caractéristiques sonores plus complexes. Par exemple, la clarté (énergie supérieure à 3000 Hz), une caractéristique acoustique connue pour corréler avec les jugements émotionnels pourrait avoir été utilisée pour discriminer les émotions (Gosselin et al., 2015 - Annexe 1; Quarto et al., 2014). Les différents niveaux de l'activité cérébrale dans le cortex auditif (Peur> Joie> Tristesse/neutralité) semblent correspondre au niveau moyen mesuré de clarté pour chaque catégorie émotionnelle (Peur> Joie> Tristesse/neutralité; tableau des différences acoustiques article 2 de cette thèse) – cette correspondance reste à être explorée expérimentalement.

L'exploration de ces différences acoustiques est d'autant plus importante pour une population ayant des troubles à les percevoir. Menant à notre troisième objectif qui était d'examiner à l’aide de simulation d’implant cochléaire comment la perception d’émotions (reconnaissance, valence, niveau d’éveil) vocales et musicales est affectée par l'implant cochléaire. Cela, afin de préciser quels indices acoustiques les utilisateurs d’IC (p.ex. le rythme et le timbre) utilisent pour discriminer les émotions vocales et musicales. Selon nos attentes, la reconnaissance d’émotions a été affectée négativement pour tous les types de stimuli (vocaux et musicaux) dans la condition de simulation d’implant cochléaire. La reconnaissance des stimuli vocaux de joie (rires) a été la plus préservée et les stimuli de peur et les stimuli musicaux exprimant la tristesse ont été les moins bien reconnus, et générallement confondus avec les stimuli joyeux ou neutres. La stratégie de recherche utilisant la simulation d’IC pourrait avoir créé un biais vers le neutre, les sons présentés à travers une simulation peuvent être perçus comme étant plus neutres par des individus ayant une audition normale, comparativement à leur perception quotidienne d’émotions auditives. Malgré cette limite, la tendance à caractériser les stimuli comme étant joyeux est similaire aux résultats observés chez les implantés cochléaires (Most & Aviner, 2009; Volkova et al., 2013; Xin Luo et al., 2007), validant ainsi l’approche de simulation.

L’examen des caractéristiques acoustiques a permis de déterminer que celles liées au timbre (la clarté, l’énergie et de la rugosité) sont utilisées pour réaliser des jugements émotionnels sous simulations d’IC. Leurs valeurs corrèlent avec les jugements d’arousal et de valence des stimuli émotionnels vocaux et musicaux. À la lumière de ces résultats, une piste prometteuse pour pallier les déficites émotionels observés chez les jeunes utilisateurs d’IC, serait un entrainement qui mettrait l’accent sur certaines propriétés acoustiques liées au timbre (la clarté, l’énergie et de la rugosité) des stimuli vocaux et musicaux.

Enseigner aux utilisateurs de IC à se concentrer sur ces propriétés acoustiques liées au timbre pourrait aider leur discrimination émotionnelle. Il est documenté que les utilisateurs IC peuvent, après un entrainement basé sur les timbres, différencier ceux-ci sur des échelles qualitatives (p.ex. de terne à brillant; Gfeller, Witt, et al., 2002). De plus grande pertinence ici, il a été observé que les utilisateurs d'IC qui suivent un entrainement (une réadaptation) musical démontrent une meilleure identification du timbre et une augmentation positive de leur évaluation subjective de la musique (Gfeller, Witt, Adamek, et al., 2002). Les bénéfices d'un tel entrainement musical sur la perception émotionnelle restent à être évalués. L'utilisation des paramètres acoustiques liés au timbre par les participants dans la condition simulation d'IC suggère qu'un entraînement à la discrimination du timbre pourrait potentiellement amélioré la reconnaissance des émotions dans les domaines musicaux et vocaux. Par conséquent, apprendre à différencier les timbres pourrait aider à discerner les émotions dans la musique et le langage. De ce fait, permettre aux jeunes utilisateurs d'implants cochléaires de développer des stratégies optimales pour favoriser leur intégration sociale et leur communication interpersonnelle.

Une des observations globales qui ressort de cette thèse est que toutes les émotions ne semblent pas traitées de manière égale par notre cerveau. Les stimuli musicaux et vocaux exprimant la peur et la joie se voient accorder un traitement préférentiel. Une activité cérébrale plus grande a été observée pour ces deux émotions stimulantes/excitantes aux extrêmes opposées de la valence émotionnelle. Lors de l'étude en IRMf (article 2 de cette thèse) l'activation cérébrale observée au niveau limbique semble indiquer un traitement préférentiel des stimuli de peur et celle observée dans le cortex auditif pour les stimuli de peur et de joie. D'un point de vue évolutif, cette préférence pourrait remonter à nos ancêtres primates non humains pour qui ces signaux servaient à avertir d'un danger ou mobiliser le groupe (Winter et al., 1966). Par exemple, la peur d'un prédateur et la joie associée à la découverte de nourriture devaient être priorisées pour assurer la survie de l'espèce. Cela dit, à travers la simulation d'implant cochléaire les participants avaient une grande difficulté à reconnaitre les stimuli de peur (vocaux et musicaux) en opposition à une facilité à identifier les stimuli joyeux - les stimuli de peur étaient notamment confondus avec les stimuli de joie). La dégradation du signal sonore (la faible résolution des fréquences sonores) ne permettrait pas à l'amygdale de jouer son rôle de détecteur de pertinence; les stimuli de peur, présentés sous simulation d'implant cochléaire ne seraient pas détectés comme pertinent/biologiquement. Ce manque de réaction de l'amygdale qui a de nombreuses connexions fonctionnelles avec le cortex auditif (Kumar, Kriegstein, & Friston, 2012; Viinikainen, Kätsyri, & Sams, 2012) pourrait nuire à la discrimination émotionnelle des stimuli très stimulants (joie, peur), par le cortex auditif sous simulation d'implant cochléaire. Cette explication pourrait entre autres expliquer le biais positif observé chez les utilisateurs IC (Hopyan et al., 2012; Stabej, Smid, Gros, Zargi, Kosir, Vatovec, et al., 2012). Pour explorer cette hypothèse, une étude en Imagerie par résonance magnétique fonctionnelle au protocole similaire à celui utilisé dans l'article 2, mais combinant la présentation de stimuli émotionnels dans le format original et sous simulation d'implant cochléaire pourait être réalisée. Celle-ci permettrait de voir si l'amygdale est impliquée dans le traitement des stimuli de peur perçus à travers une simulation d'implant cochléaire, et de déterminer si une faible résolution des fréquences sonores rend inerte aux émotions auditives l'amygdale.

3.2 Directions futures

De manière plus générale, il serait intéressant de manipuler orthogonalement les paramètres acoustiques liés aux émotions auditives (p. ex. hauteurs tonales, rythme) mais particulièrement (à la lumière de nos résultats) celles liées au timbre (la clarté, l'énergie et de la rugosité) pour observer leur impact sur la perception émotionnelle. Variée de manière continue, par exemple, la clarté permettrait de déterminer à partir de quel niveau la perception émotionnelle change. Présenter ces stimuli manipulés lors d'une étude d'imagerie permettrait de directement observer quelles caractéristiques acoustiques (ou combinaison de caractéristique acoustique) font réagir l'amygdale ou permettent la discrimination émotionnelle dans le cortex auditif. Déterminer le code combinatoire acoustique des stimuli auditifs émotionnels est, je crois, la prochaine étape pour déterminer comment les stimuli vocaux et musicaux sont traités par notre cerveau.

Une autre avenue à explorer pour comparer le traitement des émotions vocales et musicales reste l'analyse de patrons multivoxel MVPA (note de pas page, page 16). Une co-activation des régions cérébrales en IRMf est souvent interprétée comme une preuve du partage des circuits neuronaux sous-jacent. Cependant, le chevauchement d'activation ne nous fournit pas de preuves suffisantes pour parler de recyclage/partage de neurones. Les circuits établis pour la voix peuvent être adjacents à ceux utilisés pour la musique et être encore neuronalement séparables (Peretz, Vuvan, Lagrois, & Armony, 2015). L'analyse de patrons multivoxels peut être utilisée pour déterminer avec plus de précisions si des circuits neuronaux communs existent pour ces deux types de stimuli. Comme les méthodes conventionnelles, cette approche cherche à accroître la sensibilité en regardant les contributions de voxels multiples, mais sans en faire la moyenne. L'approche utilise une technique de classification des patrons d'activation pour extraire le signal qui est présent dans les activations à travers plusieurs voxels et tente de saisir les relations entre le patron spatial de l'activité IRMf et les conditions expérimentales (Mahmoudi, Takerkart, & Regragui, 2012).

Nous avons essayé d'utiliser l'analyse de patrons multivoxel, sur l'ensemble de nos données IRMf (et d'autres données pilotes). Nous avons tenté de classifier (entraîner un ordinateur à reconnaître) les émotions vocales et musicales, mais la classification supervisée n'a pas permis d'identifier les émotions au-dessus du niveau du hasard, très probablement en raison du nombre limité de stimuli. Comme spécifié dans l'introduction, la classification d'émotions avec cette technique a déjà été accomplie avec des stimuli vocaux plus longs et/ou présentés en grandes répétitions (Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009; Kotz et al., 2013).

Une étude subséquente en IRMf avec un protocole optimisé pour ce genre d'analyse (un plus grand nombre de répétitions des stimuli, des acquisitions à travers plusieurs sessions d'imageries). Cette technique est la plus prometteuse actuellement, pour permettre de déterminer avec plus de précision si les émotions vocales et musicales sont représentées à travers les mêmes patrons d'activation et de ce fait déterminer s'ils partagent les mêmes circuits neuronaux.

3.3 Implications théoriques et conclusions

La convergence d'évidence au niveau acoustique et neurophysiologique observé dans cette thèse donne un certain appui à l'hypothèse que la perception d'émotions musicales provient de leur capacité à recruter les circuits neuronaux dédiés au traitement des émotions vocales (hypothèse de recyclage neuronale; Peretz et al., 2010). Précédemment, cette hypothèse était principalement soutenue par le fait que les émotions musicales et vocales semblaient avoir des prédispositions biologiques (reconnues dès un très jeune âge, de manière universelle et de façon distincte: Peretz, 2010) et des indices acoustiques analogues (Juslin & Laukka, 2003) évoquant des jugements similaires chez les auditeurs (Ilie & Thompson, 2006).

Les résultats de cette thèse donnent un support à la possibilité d'un traitement similaire des émotions vocales et musicales dû à leur similarité allant dans le sens d'une perspective d'un traitement cérébral général, mais aussi des différences pouvant suggérer des spécificités liées à chaque domaine et/ou modalité sensorielle. La comparaison avec d'autres modes d'expressions tels que la prosodie, les visages et les gestes corporels est essentielle, elle permettra éventuellement de mieux comprendre ce traitement commun et ses limites et aiderait à définir le rôle des structures cérébrales impliquées dans le traitement émotionnel.

Le fait que la musique (une forme abstraite d'art) puisse susciter des émotions biologiquement importantes, liées à la survie humaine peut sembler paradoxal. On peut soutenir que la survie de nos ancêtres dépendait de leur capacité à détecter des codes émotionnels dans les sons, à en tirer le sens et à ajuster leurs comportements en conséquence. Il est donc fort probable que la perception d'émotions musicales provient de leur capacité à recruter les circuits neuronaux dédiés au traitement des émotions vocales; les musiciens communiqueraient les émotions aux auditeurs sur la base de principes liés à l'expression vocale émotionnelle. Il est difficile d'expliquer le contraire, les expressions émotionnelles vocales humaines sont le résultat d'expressions (d'affects) évoluées qui sont également présentes chez les primates non humains. Par conséquent, l'expression vocale est le modèle sur lequel l'expression musicale est basée plutôt que l'inverse (Juslin & Laukka, 2003). Cette affirmation ne s'applique qu'aux aspects non verbaux de la communication vocale (utilisé dans cette thèse). Il est probable que l'expression vocale des émotions se soit d'abord développée et que les expressions émotionnelles musicales se sont développées parallèlement à la parole (Brown, 2000) ou même avant la parole (Darwin, Ekman, & Prodger, 1998).

De plus, il serait logique en termes d'efficacité neuronale d'utiliser les mêmes circuits neuronaux pour le traitement des diverses émotions perçues de manière auditive. C'est sous cette perspective que Frühholz et collaborateurs (2016) proposent un modèle intégratif pour le traitement des sons émotionnels (entre autres vocaux et musicaux); un réseau de neurones commun qui facilite le décodage émotionnel à partir d'une large source de sons plutôt que des systèmes neuronaux distincts pour des types de sons émotionnels spécifiques. Leur modèle implique en son centre: l'amygdale et le cortex auditif. Le modèle attribue entre autres à l'amygdale le rôle de détecteur de pertinence (l'analyse de la valence sociale et émotionnelle des sons courts) et au cortex auditif un rôle équivalent complémentaire : le traitement des caractéristiques sonores (acoustiques) complexes. Ce qui correspond directement aux résultats observés dans notre étude en IRMf pour les stimuli vocaux et musicaux. Selon les auteurs, le traitement émotionnel auditif serait le résultat d'une communication constante entre ces deux régions centrales au modèle et leurs interactions avec les lobes frontaux, les ganglions de la base et le cervelet. Les interactions avec d'autres régions cérébrales n'ont par contre pas été corroborées par les résultats de notre étude. Peut-être que notre protocole qui impliquait que de l'écoute passive de stimuli émotionnels très courts (aucune identification / aucun jugement) n'impliquait pas d'interactions avec ses régions ou ne permettait pas l'observation d'activité dans celles-ci.

L'idée d'un système unique pour le traitement des émotions auditives est également supportée par les résultats de l'étude 3. La multitude de paramètres acoustiques utilisés par les auditeurs normaux pour réaliser des jugements émotionnels (même sous une simulation Implant cochléaire) témoigne d'une grande diversité/flexibilité dans la manière d'exprimer des émotions tant au niveau vocal que musical. Reconnaître certaines émotions dans une condition ou le signal acoustique est dégradé (simulation d'implant cochléaire, dans une foule, dans une salle ou il y a beaucoup d'écho), témoigne d'une certaine flexibilité de notre système, et probablement d'une grande redondance dans le signal acoustique nécessaire à l'identification. Plusieurs indices (partiellement redondants) donnent un système communicatif robuste qui pardonne les écarts par rapport à une utilisation optimale du ''code émotionnel vocal'' (Juslin & Laukka, 2003). Cela pourrait donc permettre au même système d'interpréter tant les émotions vocales (non verbales) que les éclats émotionnels musicaux performés au violon et à la clarinette, si celles-ci respectent en partie le ''code émotionnel auditif''.

L'utilisation d'éclats émotionnels musicaux et vocaux, dans cette thèse, a permis de rapprocher perceptuellement les expressions vocales et musicales comme il ne l'avait jamais été fait avant. En effet, aucun effet principal associé au type de stimuli n'a été observé dans l'étude d'IRMf et un traitement similaire des émotions vocales et musicales a été observé dans les structures limbiques inférieures et dans le cortex auditif. Deux études antérieures avaient directement comparé les corrélats neuronaux associés aux émotions musicales et vocales et avaient observé des régions cérébrales associées spécifiquement au traitement de la musique et de la voix et peu de résultats spécifiques liés au traitement émotionnel (Aubé et al., 2015; Escoffier et al., 2013). La grande attention portée au choix (création des stimuli) pour qu'il soit le plus comparable possible pourrait expliquer les différences observées entre les résultats de cette thèse et les leurs. Un aspect reste par contre similaire à travers les trois études, un chevauchement pour le traitement des stimuli émotionnels musical et vocal est observable le long du GTS. Cette convergence de résultats combinés à ceux de notre étude contrôlée vient attester la présence d'un circuit émotionnel commun pour les deux domaines.

L'utilisation d'éclats émotionnels musicaux est également une nouveauté de cette thèse, il avait déjà été observé que de courtes expressions (extraits de pièces musicales) étaient suffisantes pour permettre l'identification de l'émotion (Peretz et al., 1998) et même pour activer les régions limbiques comme la voix et les visages (Aubé et al., 2015). Cela suggère que les émotions musicales ne sont pas exclusivement spécifiques à la musique ou esthétiques comme l'avancent certains auteurs (Scherer, 2004). En d'autres termes, les émotions évoquées par la musique peuvent avoir un impact direct sur l'atteinte d'objectifs et la régulation émotionnelle. De plus, la musique en tant qu'outil de communication, occupe un rôle important dans les activités sociales permettant la cohésion et la coopération au sein d'un groupe (Koelsch, 2013). Ce qui renforce l'idée que les émotions musicales répondent aux critères d'émotions de base (fondamentales au bienêtre et pour la survie de l'espèce) et qu'elles sont comparables aux expressions émotionnelles vocales et celles des autres modalités sensorielles.

De plus, la création de stimuli musicaux dont l'essence était de représenter de brèves émotions musicales similaires aux éclats émotionnels vocaux non verbaux a permis d'explorer les émotions musicales d'une manière plus fondamentale que ce qui avait été précédemment fait. Elles demeurent musicales, car malgré leur courte durée, ces performances ont été réalisées par des musiciens professionnels, sur leur instrument. Leur esthétique est peut-être légèrement différente de ce qui pouvait être attendu de la musique, mais après tout, ces expressions n'étaient pas limitées par les conventions de composition occidentales – elles reflèteraient les émotions primitives musicales. Par cette définition, elles peuvent servir à étudier les mécanismes fondamentaux (comme dans cette thèse) des émotions musicales, mais ne représentent pas l'éventail d'émotions qu'il est possible d'interpréter/percevoir et qui méritent d'être étudiées. Par exemple, les émotions perçues dans la prosodie (les patrons de stress et intonation du langage) et à travers l'harmonie (les principes qui structurent la relation et la progression des accords) pourraient nécessiter un traitement spécifique à leur médium, comme le suggèrent certaines études d'imagerie (Aubé et al., 2015; Escoffier et al., 2013; Schirmer & Kotz, 2006) dans lesquelles le traitement sensoriel/émotionnel vocal et musical recrute un réseau plus étendu de régions cérébrales. Dans ces cas, en partie, leur interprétation constituerait un raffinement émotionnel spécifique à la voix ou à la musique. Les résultats présentés ont du mérite que pour décrire les mécanismes fondamentaux liés à la perception émotionnelle vocale et musicale. Mécanismes qui, selon nos résultats, seraient les mêmes pour la voix et la musique.

Bibliographie Générale

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672. http://doi.org/10.1038/372669a0

Ambert-Dahan, E., Giraud, A.-L., Sterkers, O., & Samson, S. (2015). Judgment of musical emotions after cochlear implantation in adults with progressive deafness. *Frontiers in Psychology*, *6*, 181. http://doi.org/10.3389/fpsyg.2015.00181

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A. F., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex*, *59*, 126–137. http://doi.org/10.1016/j.cortex.2014.07.013

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*(3), 399–407. http://doi.org/10.1093/scan/nsu067

Aubé, W., Peretz, I., & Armony, J. L. (2013). The effects of emotion on memory for music and vocalisations. *Memory*, *21*(8), 1–11. http://doi.org/10.1080/09658211.2013.770871

Bach, D., Schächinger, H., & Neuhoff, J. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex*, *18*(1), 145–50.

Balkwill, L.-L., & Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64. http://doi.org/10.2307/40285811

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners1. *Japanese Psychological Research*, *46*(4), 337–349. http://doi.org/10.1111/j.1468-5584.2004.00265.x

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social*, *70*(3), 614.

Behrens, G., & Green, S. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music*, *21*(1), 20–33.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. http://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. http://doi.org/10.3758/BRM.40.2.531

Bestelmeyer, P., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *The Journal of Neuroscience*, *34*(24), 8098–8105.

Bestelmeyer, P., Rouger, J., DeBruine, L., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, *117*(2), 217–223.

Bigand, E., Filipic, S., & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy*, *1060*(1), 429–437.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition &*, *19*(8), 1113–1139.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., … Murphy, D. G. M. (2011). Early Specialization for Voice and Emotion Processing in the Infant Brain. *Current Biology*, *21*(14), 1220–1224. http://doi.org/10.1016/j.cub.2011.06.009

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *98*(20), 11818–11823. http://doi.org/10.1073/pnas.191355898

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International 5.9/10*, *5*(9/10), 341–345. http://doi.org/10.1097/AUD.0b013e31821473f7

Breiter, H. C., Gollub, R. L., Weisskoff, R. M., Kennedy, D. N., Makris, N., Berke, J. D., … Hyman, S. E. (1997). Acute Effects of Cocaine on Human Brain Activity and Emotion. *Neuron*, *19*(3), 591–611. http://doi.org/10.1016/S0896-6273(00)80374-8

Brown, S. (2000). The "musilanguage" model of music evolution. In L. N. Wallin, B. Merker, & B. S. (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

Bryant, G., & Barrett, H. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*(1), 135–148.

Buchanan, T. W. T., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study, *9*(3), 227–238. http://doi.org/10.1016/S0926-6410(99)00060-9

Charlesworth, W. R., & Kreutzer, M. A. (1973). "Facial expression of infants and children." In P. Ekman (Ed.), *Darwin and Facial Expression: a Century of Research in Review* (pp. 91–168). New York, NY: Academic Press.

Cooper, J., & Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *Neuroimage*, *39*(1), 538–547.

Cottrell, S., & Mantzourani, E. (2006). The Clarinet and its players in Eastern Europe and Greece. In R. Heaton (Ed.), *The Versatile Clarinet* (pp. 31–46). New York, NY: Routledge.

Cousineau, M., Demany, L., Meyer, B., & Pressnitzer, D. (2010). What breaks a melody: Perceiving F0 and intensity sequences with a cochlear implant. *Hearing Research*, *269*(1), 34–41. http://doi.org/10.1016/j.heares.2010.07.007

Cousineau, M., McDermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *PNAS*, 109(48), 19858-19863.

Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*(3), 335–348. http://doi.org/10.1037/a0017928

Dailey, M., Cottrell, G. W., & Reilly, J. (2001). *California Facial Expressions,CAFE: Unpublished Digital Images*. San Diego, CA: University of California, Computer Science and Engineering Department.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*(7), 1019–1028. http://doi.org/10.3758/BF03194831

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), 1–10. http://doi.org/10.1016/S0010-0277(00)00136-0

Darwin, C., Ekman, P., & Prodger, P. (1998). *The expression of the emotions in man and animals*. USA: Oxford University Press.

Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. http://doi.org/10.1016/j.neuron.2007.10.004

Dellacherie, D., Hasboun, D., Baulac, M., Belin, P., & Samson, S. (2011). Impaired recognition of fear in voices and reduced anxiety after unilateral temporal lobe resection. *Neuropsychologia*, *49*(4), 618–629.

Driscoll, V. D., Oleson, J., Jiang, D., & Gfeller, K. (2009). Effects of Training on Recognition of Musical Instruments Presented through Cochlear Implant Simulations. *Journal of the American Academy of Audiology*, *20*(1), 71–82.

Dunlap, W., & Cornwell, J. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, *29*(1), 115–126.

Ekman, P. (1982). *Emotion in the Human Face*. Cambridge, MA: Cambridge University Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: Investigator's Guide*. Palo Alto: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, C. J. (2002). *Facial Action Coding System Investigator's Guide*. Salt Lake, UT: A Human Face.

Eschrich, S., Münte, T. F., Altenmüller, E. O., Panksepp, J., Eich, E., Metcalfe, J., … Halpern, A. (2008). Unforgettable film music: The role of emotion in episodic long-term memory for music. *BMC Neuroscience*, *9*(1), 48. http://doi.org/10.1186/1471-2202-9-48

Escoffier, N., Zhong, J., & Schirmer, A. (2013). Emotional expressions in voice and music: same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810.

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of Emotional Information in Voice-Sensitive Cortices. *Current Biology*, *19*(12), 1028–1033. http://doi.org/10.1016/j.cub.2009.04.054

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations, *36*(2), 480–487. http://doi.org/10.1016/j.neuroimage.2007.02.043

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts, *17*(3), 335–341. http://doi.org/10.3758/PBR.17.3.335

Flom, R., Gentile, D. A., & Pick, A. D. (2008). Infants' discrimination of happy and sad music. *Infant Behavior and Development*, *31*(4), 716–728. http://doi.org/10.1016/j.infbeh.2008.04.004

Frank, M., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social*, *80*(1), 75.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., … Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, *19*(7), 573–576. http://doi.org/10.1016/j.cub.2009.02.058

Frühholz, S., Trost, W., & Kotz, S. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96–110.

Fu, Q.-J., & Nogaki, G. (2005). Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing. *Journal of the Association for Research in Otolaryngology*, *6*(1), 19–27. http://doi.org/10.1007/s10162-004-5024-3

Gabrielsson, A., & Juslin, P. N. (1996). Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music*, *24*(1), 68–91. http://doi.org/10.1177/0305735696241007

Gabrielsson, A., & Juslin, P. N. (2003). *Emotional expression in music*. Oxford University Press.

Gabrielsson, A., & Lindström, E. (2001). *The influence of musical structure on emotional expression*. Oxford University Press.

Gfeller, K., & Lansing, C. (1992). Musical Perception of Cochlear Implant Users as Measured by the Primary Measures of Music Audiation: An Item Analysis. *Journal of Music Therapy*, *29*(1), 18–39. http://doi.org/10.1093/jmt/29.1.18

Gfeller, K., & Lansing, C. R. (1991). Melodic, Rhythmic, and Timbral Perception of Adult Cochlear Implant Users. *Journal of Speech Language and Hearing Research*, *34*(4), 916. http://doi.org/10.1044/jshr.3404.916

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., … Stordahl, J. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International*, *3*(1), 29–53. http://doi.org/10.1002/cii.50

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., & Ringgenberg, S. (2002). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology*, *13*(3), 132–45.

Gfeller, K., Witt, S., Mehr, M., & Woodworth, G. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Annals of Otology, Rhinology & Laryngology*, *111*(4), 349–356.

Gilbers, S., Fuller, C., Gilbers, D., Broersma, M., Goudbeek, M., Free, R., & Başkent, D. (2015). Normal-Hearing Listeners' and Cochlear Implant Users' Perception of Pitch Cues in Emotional Speech. *I-Perception*, *6*(5), 301006615599139. http://doi.org/10.1177/0301006615599139

Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex*, *71*, 171–182. http://doi.org/10.1016/j.cortex.2015.06.022

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244. http://doi.org/10.1016/j.neuropsychologia.2006.07.012

Gosselin, N., Peretz, I., Noulhiane, M., & Hasboun, D. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, *128*(3), 628–640.

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., & Samson, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain : A Journal of Neurology*, *128*(Pt 3), 628–40. http://doi.org/10.1093/brain/awh420

Green, A. C., Bærentsen, K. B., Stødkilde-Jørgensen, H., Wallentin, M., Roepstorff, A., & Vuust, P. (2008). Music in minor activates limbic structures: a relationship with dissonance? *NeuroReport*, *19*(7), 711–715. http://doi.org/10.1097/WNR.0b013e3282fd0dd8

Grossmann, T., Oberecker, R., Koch, S., & Friederici, A. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*(6), 852–858.

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2141–2155.

Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). Worth a thousand words: Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*(3), 293–305. http://doi.org/10.1037/a0015178

Heilman, K. M., Scholes, R., & Watson, R. T. (1975). Auditory affective agnosia. Disturbed comprehension of affective speech. *Journal of Neurology, Neurosurgery, and Psychiatry*, *38*(1), 69–72. http://doi.org/10.1136/JNNP.38.1.69

Hopyan-Misakyan, T., Gordon, K., & Dennis, M. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, *15*(2), 136–46.

Hopyan, T., Manno III, F. A., Papsin, B. C., & Gordon, K. A. (2016). Sad and happy emotion discrimination in music by children with cochlear implants. *Child Neuropsychology*, *22*(3), 366–380. http://doi.org/10.1080/09297049.2014.992400

Hopyan, T., Peretz, I., Chan, L. P., Papsin, B. C., & Gordon, K. A. (2012). Children Using Cochlear Implants Capitalize on Acoustical Hearing for Music Perception. *Frontiers in Psychology*, *3*, 425. http://doi.org/10.3389/fpsyg.2012.00425

Hyde, K. L., & Peretz, I. (2004). Brains That Are out of Tune but in Time. *Psychological Science*, *15*(5), 356–360. http://doi.org/10.1111/j.0956-7976.2004.00683.x

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception*, *23*(4), 319–330. http://doi.org/10.1525/mp.2006.23.4.319

Johnson-Laird, P., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition & Emotion*, *6*(3–4), 201–223.

Juslin, P., & Laukka, P. (2003). Emotional expression in speech and music. *Annals of the New York Academy of*, *1000*(1), 279–282.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1812. http://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N. (2012). Are Musical Emotions Invariant Across Cultures? *Emotion Review*, *4*(3), 283–284. http://doi.org/10.1177/1754073912439773

Juslin, P. N. (Ed), & Sloboda, J. A. (Ed). (2001). *Music and emotion: Theory and research.* Oxford University Press.

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion*, *8*(5), 668–683. http://doi.org/10.1037/a0013505

Juslin, P. N. P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. http://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575. http://doi.org/10.1017/S0140525X08005293

Kantrowitz, J., Leitman, D., Lehrfeld, J., & Laukka, P. (2011). Reduction in tonal discriminations predicts receptive emotion processing deficits in schizophrenia and schizoaffective disorder. *Schizophrenia Bulletin*, *39*(1), 86–93.

Khalfa, S., Peretz, I., Blondin, J.-P., & Robert, M. (2002). *Event-related skin conductance responses to musical emotions in humans*. *Neuroscience Letters* (Vol. 328). http://doi.org/10.1016/S0304-3940(02)00462-7

Khalfa, S., Schon, D., Anton, J., & Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport*, *16*(18), 1981–1984.

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, S14. http://doi.org/10.1068/v070821

Knutson, B., & Cooper, J. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*(4), 411–417.

Koeda, M., Belin, P., Hama, T., Masuda, T., & Matsuura, M. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology*, *4*(105).

Koelsch, S. (2013). Emotion and Music. In: J.L., Armony, & P. Vuilleumier (eds.). *The Cambridge handbook of human affective neuroscience*. Cambridge. New York: Cambridge University Press. pp. 286-303.

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy, *81*, 49–60. http://doi.org/10.1016/j.neuroimage.2013.05.008

Kong, Y.-Y., Cruz, R., Jones, J. A., & Zeng, F.-G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing*, *25*(2), 173–85.

Kotz, S., Kalberlah, C., & Bahlmann, J. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain*, *34*(8), 1971–1981.

Kotz, S., Meyer, M., Alter, K., Besson, M., & Cramon, D. von. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain and Language*, *86*(3), 366–76.

Kumar, S., Kriegstein, K. von, & Friston, K. (2012). Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds. *The Journal of*.

Laneau, J., Wouters, J., & Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *Journal of the Acoustical*, 3606–3619.

Lang, P. J., Öhman, A., & Vaitl, D. (1988). *The International Affective Picture System*. Gainesville, FL: University of Florida, Center for Research in Psychology.

Language, T., & Computing, T. (2004). MATLAB The Language of Technical Computing. *Components*, *3*(7), 750. http://doi.org/10.1007/s10766-008-0082-5

Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*, 1–8.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145. http://doi.org/10.1016/j.cub.2010.12.033

Laukka, P., Eerola, T., Thingujam, N. N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449. http://doi.org/10.1037/a0031388

Laukka, P., & Juslin, P. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, *31*(3), 182–191.

Lehmann, A., & Paquette, S. (2015). Cross-domain processing of musical and vocal emotions in cochlear implant users. *Frontiers in Neuroscience*, *9*, 343. http://doi.org/10.3389/fnins.2015.00343

Lima, C. F., & Castro, S. L. (2011). Emotion recognition in music changes across the adult life span. *Cognition and Emotion*, *25*(4), 585–98.

Lima, C., Garrett, C., & Castro, S. (2013). Not all sounds sound the same: Parkinson's disease affects differently emotion processing in music and in speech prosody. *Journal of Clinical and Experimental*, *35*(4), 373–92.

Limb, C. J. (2006). Cochlear implant-mediated perception of music. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *14*(5), 337–340. http://doi.org/10.1097/01.moo.0000244192.59184.bd

Longfellow, H. (1857). *Outre-mer: a pilgrammage beyond the sea*. Routledge & Co.

Lonsdale, A. J. A., & North, A. C. A. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, *102*(1), 108–134. http://doi.org/10.1348/000712610X506831

Looi, V., Gfeller, K., & Driscoll, V. (2012). Music Appreciation and Training for Cochlear Implant Recipients: A Review. *Seminars in Hearing*, *33*(4), 307–334. http://doi.org/10.1055/s-0032-1329222

Mahmoudi, A., Takerkart, S., & Regragui, F. (2012). Multivoxel pattern analysis for fMRI data: A review. *Methods in Medicine*, *2012*.

Mas-Herrero, E., Zatorre, R., & Rodriguez-Fornells, A. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology*, *24*(6), 699–704.

Menon, V., & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *NeuroImage*, *28*(1), 175–184. http://doi.org/10.1016/j.neuroimage.2005.05.053

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410–1421. http://doi.org/10.1016/S0028-3932(03)00017-4

Mitterschiffthaler, M. M. T., Fu, C. H. Y. C., Dalton, J. A. J., Andrew, C. M., & Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, *28*(11), 1150–1162. http://doi.org/10.1002/hbm.20337

Monrad-Krohn, G. H. (1963). “The third element of speech: prosody and its disorders.” In L. Halpern (Ed.), *Problems of Dynamic Neurology* (pp. 101–107). Jerusalem: Hebrew University Press.

Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, *114*(1), 408. http://doi.org/10.1121/1.1577552

Morris, J. ., Scott, S. ., & Dolan, R. . (1999). Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia*, *37*(10), 1155–1163.

http://doi.org/10.1016/S0028-3932(99)00015-9

Most, T., & Aviner, C. (2009). Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing AIDS, and normal hearing. *Journal of Deaf Studies and Deaf Education*, *14*(4), 449–64. http://doi.org/10.1093/deafed/enp007

Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, *131*(2), 1307. http://doi.org/10.1121/1.3672697

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–99.

Nogaki, G., Fu, Q.-J., & Galvin, J. J. (2007). Effect of training rate on recognition of spectrally shifted speech. *Ear and Hearing*, *28*(2), 132–40. http://doi.org/10.1097/AUD.0b013e3180312669

Norman, K., Polyn, S., Detre, G., & Haxby, J. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430.

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of Music in Everyday Life. *Music Perception: An Interdisciplinary Journal*, *22*(1), 41–77.

Omar, R., Henley, S., Bartlett, J., & Hailstone, J. (2011). The structural neuroanatomy of music emotion recognition: evidence from frontotemporal lobar degeneration. *Neuroimage*, *56*(3), 1814–21.

Paquette, S., Ahmed, D., Peretz, I., & Lehmann, A. (n.d.). Recognition of musical and vocal emotions through cochlear implants simulation. *Neuropsychologia*.

Paquette, S., Peretz, I., & Belin, P. (2013). The Musical Emotional Bursts: a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology*, *4*, 509. http://doi.org/10.3389/fpsyg.2013.00509

Pell, M. (2006). Judging emotion and attitudes from prosody following brain damage. *Progress in Brain Research*, *156*, 303–317.

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. http://doi.org/10.1016/j.biopsycho.2015.08.008

Pell, M., & Kotz, S. (2011). On the time course of vocal emotion recognition. *PLoS One*, *6*(11).

Pell, M., Monetta, L., Paulmann, S., & Kotz, S. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107–120.

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., … Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(10), 4758–63. http://doi.org/10.1073/pnas.0909074107

Peretz, I. (2010). Towards a Neurobiology of Musical Emotions. In J. Juslin, P; Sloboda (Ed.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 99–126). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199230143.003.0005

Peretz, I., Aubé, W., & Armony, J. (2013). Towards a neurobiology of musical emotions. In E. Altenmüller, S. Schmidt, & E. Zimmerman (Eds.), *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*. Oxford: Oxford University Press.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., & B, J. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron*, *17*(33), 185–91.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. http://doi.org/10.1016/S0010-0277(98)00043-2

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140090. http://doi.org/10.1098/rstb.2014.0090

Peretz, I., & Zatorre, R. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89–114.

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in psychology*, 3, 23.

Poissant, S. F., Whitmal, N. A., & Freyman, R. L. (2006). Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. *The Journal of the*

Acoustical Society of America, *119*(3), 1606. http://doi.org/10.1121/1.2168428

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of Facial Expressions and Voices and of their Combination in the Human Brain. *Cortex*, *41*(1), 49–59. http://doi.org/10.1016/S0010-9452(08)70177-1

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114*(1), 446. http://doi.org/10.1121/1.1579009

Quarto, T., Blasi, G., Pallesen, K. J., Bertolino, A., & Brattico, E. (2014). Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits. *PloS One*, *9*(7), e103278. http://doi.org/10.1371/journal.pone.0103278

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(10), 3843–52. http://doi.org/10.1523/JNEUROSCI.4515-10.2011

Rothauser, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Nordby, K. S., Silbiger, H. R., … Weinstock, M. (1969). IEEE Recommnded Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246. http://doi.org/10.1109/TAU.1969.1162058

Roy, M., Mailhot, J.-P., Gosselin, N., Paquette, S., & Peretz, I. (2009). Modulation of the startle reflex by pleasant and unpleasant music. *International Journal of Psychophysiology*, *71*(1), 37–42. http://doi.org/10.1016/j.ijpsycho.2008.07.010

Royet, J. P., Zald, D., Versace, R., Costes, N., Lavenne, F., Koenig, O., & Gervais, R. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(20), 7752–9.

Salimpoor, V. V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature*, *14*(2), 257–262. http://doi.org/10.1038/nn.2726

Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala: An Evolved System for Relevance Detection. *Reviews in the Neurosciences*, *14*(4), 303–313. http://doi.org/10.1515/REVNEURO.2003.14.4.303

Sander, K., Brechmann, A., & Scheich, H. (2003). Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Research Protocols*, *11*(2), 81–91. http://doi.org/10.1016/S1385-299X(03)00018-7

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408–2412. http://doi.org/10.1073/pnas.0908239106

Sauter, D., Eisner, F., Ekman, P., & Scott, S. (2015). Emotional vocalizations are recognized across cultures regardless of the valence of distractors. *Psychological Science*, *26*(3), 354–56.

Scherer, K. (1994). Affect bursts: Essays on emotion theory. In S. H. M. van Goozen, N. E. van de Poll, & Sergeant J. A. (Eds.), *Emotions: Essays on emotion theory* (pp. 161–193). Hillsdale, NJ: Lawrence Erlbaum.

Scherer, K. (1995). Expression of emotion in voice and music. *Journal of Voice*.

Scherer, K., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural*, *32*(1), 76–92.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. http://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2004). Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them? *Journal of New Music Research*, 33(3), 239–251.

Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *NeuroImage*, *63*(1), 137–147. http://doi.org/10.1016/j.neuroimage.2012.06.025

Schirmer, A., & Kotz, S. A. S. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. http://doi.org/10.1016/j.tics.2005.11.009

Schirmer, A., Striano, T., & Friederici, A. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635–639.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication*, *40*(1), 99–116.

Scott, S., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997).

Auditory recognition of emotion after amygdalotomy: impairment of fear and anger. *Nature*, *385*, 254–257.

Sharda, M., Midha, R., Malik, S., Mukerji, S., & Singh, N. C. (2015). Fronto-Temporal Connectivity is Preserved During Sung but Not Spoken Word Listening, Across the Autism Spectrum. *Autism Research*, *8*(2), 174–186. http://doi.org/10.1002/aur.1437

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., & Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain : A Journal of Neurology*, *124*(9), 1720–33. http://doi.org/10.1093/brain/124.9.1720

Spencer, H. (1857). The origin and function of music. *Essays, Scientific, Political, and Speculative*, *2*.

Sprengelmeyer, R., Rausch, M., Eysel, U. T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings. Biological Sciences / The Royal Society*, *265*(1409), 1927–31. http://doi.org/10.1098/rspb.1998.0522

Sprengelmeyer, R., Young, A. W., Schroeder, U., Grossenbacher, P. G., Federlein, J., Buttner, T., & Przuntek, H. (1999). Knowing no fear, *266*(1437). http://doi.org/10.1098/rspb.1999.0945

Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., & Vatovec, J. (2012). The music perception abilities of prelingually deaf children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *76*(10), 1392–1400. http://doi.org/10.1016/j.ijporl.2012.07.004

Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., Vatovec, J., … al., et. (2012). The music perception abilities of prelingually deaf children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *76*(10), 1392–400. http://doi.org/10.1016/j.ijporl.2012.07.004

Terwogt, M. M., & Van Grinsven, F. (1991). Musical Expression of Moodstates. *Psychology of Music*, *19*(2), 99–109. http://doi.org/10.1177/0305735691192001

Thompson, W., & Balkwill, L. (2006). Decoding speech prosody in five languages. *Semiotica*, *2006*(158), 407–424.

Thompson, W. F. W., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *PNAS*, *109*(46), 19027–19032. http://doi.org/10.1073/pnas.1210344109

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752.

Viinikainen, M., Kätsyri, J., & Sams, M. (2012). Representation of perceived sound valence in the human brain. *Human Brain Mapping*, *33*(10), 2295–2305. http://doi.org/10.1002/hbm.21362

Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., & Gordon, K. A. (2013). Children with bilateral cochlear implants identify emotion in speech and music. *http://dx.doi.org/10.1179/1754762812Y.0000000004*, *14*(2), 80–91.

Wang, D. J., Trehub, S. E., Volkova, A., & van Lieshout, P. (2013). Child implant users' imitation of happy- and sad-sounding speech. *Frontiers in Psychology*, *4*(351). http://doi.org/10.3389/fpsyg.2013.00351

Wang, Y., Su, Y., Fang, P., & Zhou, Q. (2011). Facial expression recognition: Can preschoolers with cochlear implants and hearing aids catch it? *Research in Developmental Disabilities*, *32*(6), 2583–2588.

Wiefferink, C. H., Rieffe, C., Ketelaar, L., De Raeve, L., & Frijns, J. H. M. (2013). Emotion understanding in deaf children with a cochlear implant. *Journal of Deaf Studies and Deaf Education*, *18*(2), 175–86. http://doi.org/10.1093/deafed/ens042

Wiefferink, C. H., Rieffe, C., Ketelaar, L., & Frijns, J. H. M. (2012). Predicting social functioning in children with a cochlear implant and in normal-hearing children: The role of emotion regulation. *International Journal of Pediatric Otorhinolaryngology*, *76*(6), 883–889. http://doi.org/10.1016/j.ijporl.2012.02.065

Winter, P., Ploog, D., & Latta, J. (1966). Vocal repertoire of the squirrel monkey (Saimiri sciureus), its analysis and significance. *Experimental Brain Research*, *1*(4), 359–384.

Xin Luo, X., Fu, Q.-J., & Galvin, J. J. (2007). Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users. *Trends in Amplification*, *11*(4), 301–315. http://doi.org/10.1177/1084713807305301

Yik, M., Russell, J., & Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social*, *77*(3), 600.

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory

Cortex. *Cerebral Cortex*, *11*(10), 946–953. http://doi.org/10.1093/cercor/11.10.946

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. http://doi.org/10.1037/1528-3542.8.4.494

Annexe 1. Sensitivity to musical emotions in congenital amusia

Publié dans Cortex (2015), volume 71, pages 171-82.

Nathalie Gosselin[1], Sébastien Paquette[1], Isabelle peretz[1]
*All authors contributed equally to the work.

[1] Department of Psychology, International Laboratory for Brain Music and Sound Research, Center for Research on Brain, Language and Music, University of Montreal, Montreal, QC, Canada

Keywords: Amusia, Music, Tone deafness, Emotion recognition, Musical mode, Tempo, Pitch

Sensitivity to musical emotions in congenital amusia

Abstract

The emotional experience elicited by music is largely dependent on structural characteristics such as pitch, rhythm, and dynamics. We examine here to what extent amusic adults, who have experienced pitch perception difficulties all their lives, still maintain some ability to perceive emotions from music. Amusic and control participants judged the emotions expressed by unfamiliar musical clips intended to convey happiness, sadness, fear and peacefulness (Experiment 1A). Surprisingly, most amusic individuals showed normal recognition of the four emotions tested here. This preserved ability was not due to some peculiarities of the music, since the amusic individuals showed a typical deficit in perceiving pitch violations intentionally inserted in the same clips (Experiment 1B). In Experiment 2, we tested the use of two major structural determinants of musical emotions: tempo and mode. Neutralization of tempo had the same effect on both amusics' and controls' emotional ratings. In contrast, amusics did not respond to a change of mode as markedly as controls did. Moreover, unlike the control participants, amusics' judgments were not influenced by subtle differences in pitch, such as the number of semitones changed by the mode manipulation. Instead, amusics showed normal sensitivity to fluctuations in energy, to pulse clarity, and to timbre differences, such as roughness. Amusics even showed sensitivity to key clarity and to large mean pitch differences in distinguishing happy from sad music. Thus, the pitch perception deficit experienced by amusic adults had only mild consequences on emotional judgments. In sum, emotional responses to music may be possible in this condition.

Introduction

The most common motive for listening to music is its rich emotional content (Sloboda & O'Neill, 2001). Such an emotional appeal may not resonate for those who have a lifelong disorder in processing the pitch structure of music. The disorder, known as congenital amusia, is typically expressed by poor musical pitch perception and impoverished musical experiences in an otherwise normally developing system (Peretz, 2013). In particular, individuals with amusia (amusics hereafter) can hardly detect a semitone difference, which is the smallest pitch interval and the building block of everyday music (Hyde & Peretz, 2004; Vuvan, Nunes-Silva, & Peretz, 2015). As a result, amusics are indifferent to dissonance (Ayotte, Peretz, & Hyde, 2002; Cousineau, McDermott, & Peretz, 2012). They should also show little differentiation between the major and minor modes that underlie the happysad distinction in Western musical culture. From a musictheoretic perspective, both dissonance and mode discrimination depend on a semitone difference.

Thus, a central question in the study of congenital amusia is to what extent the perception of musical emotions is affected by amusics' poor discrimination of semitone differences. Indeed, the diagnosis of congenital amusia is dependent on the inability to discriminate melodies that differ by a semitone or more. In general, an amusia diagnosis rests on a global score lying two standard deviations below the mean of the normal population on a battery of six tests, the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003). These tests require same/different discrimination of melodies that may contain single-note changes in pitch or duration, metrical judgments (i.e., if a melody is a march or waltz), and the recognition of novel melodies against those previously presented. Under standard use, the MBEA does not cover emotion discrimination in music. A prior attempt to do so showed that the emotion discrimination subtest used in the MBEA lacked sensitivity: amusics and controls both performed at ceiling (Sloboda, Wise, & Peretz, 2005). Thus, the question of whether the ability to recognize emotions from music is retained in cases of congenital amusia remains open.

Emotion recognition has both clinical and neurological relevance since congenital amusia is a neurogenetic disorder. It is characterized by impoverished connectivity between the auditory cortex and the inferior frontal gyrus of the right hemisphere (Albouy et al., 2013; Hyde et al., 2006, 2007; Hyde, Zatorre, & Peretz, 2011; Loui, Alsop, & Schlaug, 2009). A brain marker of this poor connectivity is the absence of normal late positivities (P300/P600) when detecting a pitch deviation that is smaller than a semitone in a simple tone sequence (Moreau, Jolicœur, & Peretz, 2013; Peretz, Brattico, & Tervaniemi, 2005) or that is mistuned in a melody (Peretz, Brattico, Järvenpää, & Tervaniemi, 2009; Zendel, Lagrois, Robitaille, & Peretz, 2015). This electrical marker of congenital amusia can be observed even in children (Lebrun, Moreau, McNally-Gagnon, Mignault Goulet, & Peretz, 2012). Such biomarkers consistently accompany the congenital amusia phenotype, which is characterized by difficulty detecting out-of-key notes in melodies (Peretz, Cummings, & Dubé, 2007). The phenotype is found in 39% of first-degree relatives of amusic individuals, whereas only 3% have such a deficit in the control families (Peretz et al., 2007). The family aggregation pattern of transmission coupled with the observation of reduced connectivity and thicker cortex (Hyde et al., 2007) suggest the presence of defects in neuronal migration that can be caused by a single gene mutation.

However, music abilities, as with most complex cognitive abilities, owe their ultimate functional properties not only to genetic factors but also to experience-based plasticity. One important factor that could influence plasticity in amusia is the degree of enjoyment amusics can get from listening to music, and therefore, their level of engagement with music. Amusics vary widely in their musical interest. A few report irritation, many feel indifferent, and a minority are music lovers (Ayotte et al., 2002; McDonald & Stewart, 2008; Omigie et al., 2015). Curiously, this variability in music interest appears unrelated to the severity of the disorder (McDonald & Stewart, 2008). These observations may reflect dissociations between emotional and cognitive processing of music. For example, a lack of interest in music, termed musical anhedonia, can occur in individuals who have both normal music perception abilities and normal music emotion recognition (Mas-Herrero, Zatorre, & Rodriguez-Fornells, 2014). Another intriguing dissociation between emotion and perception can be observed after brain damage. For example, I.R., a case we have described previously, enjoyed listening to music and was normal at discriminating happy from sad music, but was unable to recognize those same musical excerpts (Peretz & Gagnon, 1999; Peretz, Gagnon, & Bouchard, 1998). The possibility of finding patterns of dissociation between perception, emotions, and appreciation in music processing provided an incentive for a detailed assessment of music emotion recognition in congenital amusia.

To this aim, in the present study we assessed groups of amusics and matched controls with the same tests of musical emotions that we have used in brain-damaged populations in prior studies (e.g. Gosselin et al., 2005; Peretz et al., 1998). For example, the emotional clips (Experiment 1A) have been validated in the normal adult population (Vieillard et al., 2008) and have been shown to be sensitive to the resection of the medial temporal lobe (Gosselin et al., 2005; Gosselin, Peretz, Hasboun, Baulac, & Samson, 2011) and to damage to the amygdala (Gosselin, Peretz, Johnsen, & Adolphs, 2007). The test, unlike the emotional test once used in the MBEA, does not present ceiling effects and is sensitive to the presence of a selective deficit. The test probes the recognition of four emotions: happiness, sadness, peacefulness and fear. Since amusics are able to use the temporal dimension in music discrimination to some extent, they may be able to use tempo (corresponding to the speed or pace of a given piece) and probably other non-pitch based cues such as dynamics, articulation and timbre to guide their perceptual judgments (Gabrielsson & Lindström, 2010). Accordingly, they are expected to be only mildly impaired in the identification of the emotions conveyed in the music presented in Experiment 1A. In contrast, amusics are expected to fail to detect out-of-key notes in those same stimuli. This is assessed in Experiment 1B.

In order to test the contribution of mode to amusics' perception of musical emotions, another validated set of musical excerpts was used in Experiment 2. Using this set with I.R., the case of acquired amusia mentioned above, we observed that emotional judgments based on either tempo or mode can be intact despite the presence of severe music perception and memory deficits (Peretz et al., 1998). However, I.R. was also able to perceive pitch differences as small as a semitone (Peretz, Blood, Penhune, & Zatorre, 2001). In contrast, congenital cases of amusia have difficulties detecting a semitone change in pitch, as described previously. Thus, musical mode was expected to be a poor structural cue for amusics. We predicted that amusics might recognize emotions from music to some extent when other structural characteristics, such as tempo, are available, but would have difficulties when mode is the principal determinant.

Method

Participant

In total, 13 amusics (of which 10 took part in all experiments) and 15 controls (of which eight completed all experiments) participated (Table 1). Amusics and controls were matched for age, number of years of education, and years of musical training. The only known difference between the two groups was that each amusic individual performed more than two standard deviations below both the global score and the melodic composite score obtained by controls on the MBEA (with 22 and 20.3/30, respectively, as cut-off scores adjusted for age and education). Individual performance on the pitch change detection task (Hyde & Peretz, 2004) is also provided in Table 1. While amusics differed in the severity of their impairment in the pitch change detection task, all were impaired, with scores at least 2 SDs below controls' mean.

To get a better grasp of the subjective experience of musical emotions, we asked 12 (all but A11) amusics and 14 controls to complete the Affective Responses to Music questionnaire used by McDonald and Stewart (2008). This questionnaire includes 10 questions (Appendix 1), of which 5 are similar to those used by Mas-Herrero, Marco-Pallares, Lorenzo-Seva, Zatorre, and Rodrıguez-Fornells (2013) in their assessment of anhedonia. Participants were asked to indicate whether they agreed with each statement, such as “certain music can put me in a good mood or match my current mood”, by selecting (1) strongly agree, (2) agree, (3) unsure, (4) disagree, or (5) strongly disagree. We computed the average of the 10 ratings provided by each participant. The higher the average score, the more the participant was emotionally disengaged or anhedonic. The amusics' scores ranged from 1.2 to 4.2 (Table 1) and confirmed the heterogeneity of selfreports. This variability may explain why their scores did not differ reliably from controls' scores [range 1-3.1; $t(25) = 1.89$, $p = .07$]. When considering a score lying 2 SD above the normals' mean (that is, above 2.8) as indicative of anhedonia, only two amusics (A6 and A7) and one control qualified as anhedonic.

[Insert Table 1 here]

All participants provided informed consent and the protocol was approved by the Ethics committee of the University of Montreal.

Experiment 1A: recognition of happy, sad, scary and peaceful music

In order to measure the ability to recognize musical emotions, we presented a validated set of musical clips expressing happiness, sadness, fear, or peacefulness (Vieillard et al., 2008) to amusics and matched controls. Their task was to judge to what extent each clip expressed each of the four emotions, allowing mixed emotions and ambiguity to be expressed in the ratings.

Participants

Twelve amusics (A1 to A12) and 12 matched controls participated in this first experiment.

Material

The full set of 56 musical clips was used here (see Vieillard et al., 2008, for a detailed description and validation). The clips were composed for research purposes in the genre of movie soundtracks and followed the rules of the Western tonal system. Each clip contains a melody with accompaniment and has a regular rhythm, with the exception of seven scary excerpts that were judged to have irregular rhythms in a previous study (Gosselin et al., 2005). The happy excerpts are in a major mode with a relatively fast tempo (137 beat per minute, BPM; range 92-196). In contrast, the sad excerpts are in a minor mode at a slow tempo (46 BPM; range 40-60). The peaceful music is composed in a major mode with an intermediate tempo (74 BPM; range 54-100), while the scary music (fear) is more variable (range 44-172 BPM) and composed with minor chords. The mean duration of the clips is 12.4 sec (range 9.2-16.4), and the duration is matched across the four emotion categories. The clips are computer-generated with a piano timbre. A full description, including the music notation, can be found at www.brams.umontreal.ca/short/emotional_ clips.

The musical clips were further analyzed and quantified for the presence of various acoustical features known to affect judgments (Quarto, Blasi, Pallesen, Bertolino, & Brattico, 2014). The values are presented in Table 2. To measure the mean pitch, we averaged all of the notes in MIDI where each number from 0 to 127 corresponds to a semitone. The lowest note (A0) of a standard piano is 21 in MIDI and the highest (C8) is 108. These values were computed with the MIDIToolbox (Eerola & Toiviainen, 2004). Values for key and pulse clarity, as well as timbre, defined by brightness or amount of energy above 3000 Hz (Juslin, 2000), root mean square (RMS) energy, and roughness (beating), were extracted with the MIRToolbox 1.5 (Lartillot & Toiviainen, 2007). Note that there are slight variations in RMS energy within each melody, which are captured by the MIR toolbox, even though the stimuli were normalized using the maximum peak value of each stimulus.

The musical clips can be distinguished on the basis of all considered acoustical characteristics except brightness, as assessed with ANOVA and Bonferroni correction for multiple comparisons. For the mean pitch, a main effect of Emotion (happy, sad, scary, and peaceful) was observed, F(3, 52) = 6.30, p < .002, $\eta^2_{partial}$ =.27. The happy musical stimuli were on average seven semitones higher than the sad stimuli and eleven semitones higher than the scary stimuli, t(26) = 5.59 and 3.18, respectively, both p < .004. Similarly, a main effect of Emotion was observed for key clarity, F(3, 52) = 5.85, p < .003, $\eta^2_{partial}$ =.25, and pulse clarity, F(3, 52) = 18.54, p < .001, $\eta^2_{partial}$ = .52. Both pulse and key were less clear in the scary clips than in the happy and peaceful clips (all p < .009). Pulse clarity was also less clear in the scary excerpts compared to the sad ones, t(26) = 4.25, p < .001. Finally, the clips differed on at least two timbral qualities: RMS energy, F(3, 52) = 8.59, p < .001, $\eta^2_{partial}$ = .33, and roughness, F(3, 52) = 5.17, p < .004, $\eta^2_{partial}$ = .23. For both characteristics, the happy stimuli obtained higher values than the sad and peaceful stimuli (p < .008). These differences in timbral qualities may be related to the higher number of tones in the happy clips. Brightness did not reach significance, F(3, 52) = 1.96, p = .131, $\eta^2_{partial}$ = .10.

[Insert Table 2 here]

Procedure

The clips were presented in one of two random orders (as in previous validation studies; Vieillard et al., 2008; Gosselin et al., 2005). They were played at a comfortable listening level through BeyerDynamic DT 990 Pro headphones in a sound-attenuated booth. Each participant was tested individually and asked to judge to what extent each clip expressed each of the four emotions (happiness, sadness, fear and peacefulness) using a 10-point scale, from 0 = absent to 9 = present. Specifically, each participant recorded four ratings for each clip, one for each emotion, since each musical clip could potentially express more than one emotion. No feedback regarding the emotion intention was given, with the exception of the two practice examples (one intended to convey sadness and one fear). The experiment lasted approximately 45 min.

Results and comments

For each labeled emotion, we calculated the mean rating given by each participant. Although there was an order effect, $F(1, 20) = 4.70$, $p = .042$, $\eta^2_{partial} = .19$, it did not interact with the factors of interest (Order x Group: F<1; Order x Emotion: F<1). Therefore, the ratings were averaged across the two orders. As shown in Table 3, participants gave the highest rating for the emotional label that corresponded to the intended emotion. Amusics' and controls' ratings overlapped for the happy and sad clips (Fig. 1). However, a few amusics rated peaceful and scary stimuli lower than controls (peaceful: A6 and A11; scary: A2 and A6).

The ratings were analyzed by a two-way mixed-design ANOVA considering Group (amusic vs control) as the betweensubjects factor and Emotion (happiness, sadness, fear, and peacefulness) as the within-subject factor. The analysis revealed no significant effect of Group $F(1, 22) = 3.19$, $p = .09$, $\eta^2_{partial} = .13$, nor significant interaction between Group and Emotion, $F(3, 55) = 1.47$, $p = .24$, $\eta^2_{partial} = .06$ (with Greenhouse-Geisser correction factor, 3= .84). While the possibility that larger group sizes would have led to a signifi- cant result cannot be discounted, the musical impairments that characterize congenital amusia do not seem to have a large impact on emotional judgments at the group level.

The main effect of Emotion was significant, $F(3, 55) = 29.53$, $p < .001$, $\eta^2_{partial} = .57$ (with Greenhouse-Geisser correction factor, 3= .84). Happiness was given higher scores on the 0 to 9 scale indicating how much each clip expressed happiness, when compared to how much the sad, scary, and peaceful music was thought to express sadness, fear, and peacefulness, respectively, $t(26) = 7.30, 8.93, 6.99$, for the comparison of happy with sad, scary and peaceful music, respectively, all $p < .008$ with Bonferroni correction. Sad music was also given higher scores than scary music, $t(26) = 3.68$, $p < .002$. The scary clips were generally given low scores across all emotions, which could be due to the older age of the participants (Lima & Castro, 2011).

[Insert Table 3 here] [Insert Figure 1 here]

Finally, we did not find relationships between the ratings and the acoustical characteristics considered in Table 2. The only correlation to reach significance concerned the ratings obtained for the peaceful music and their RMS energy values, $r(12) = .59$, $p < .03$.

We also computed correlations between participants' pitch discrimination abilities (using the scores obtained in the pitch change detection task; see Table 1) and the ratings provided for each intended emotion. The only correlation to reach signifi- cance was obtained in amusics for the scary clips [$r(9) = .60$, $p = .047$]. In this case, the larger the pitch discrimination deficit, the lower the ratings of fear expressed in the clips.

Experiment 1B: error detection in emotional clips

The near-normal performance of the amusic group in musical emotion discrimination raised the possibility that the clips were easier to process than the stimuli usually used for the diagnosis and the characterization of amusia (that is, the melodies of the MBEA). To assess this possibility, we presented here the same musical clips used in Experiment 1A, but in the form of an error detection task, as in our prior studies of acquired amusia (Peretz et al., 1998) and of temporal lobe excision (Gosselin et al., 2005). In the error detection task, a semitone pitch-shift was applied to one entire measure of the melody while the accompaniment remained unchanged for half of the clips presented. This manipulation creates local (sensory) dissonance, which is typically difficult for amusics to detect, whereas it is easily picked up by controls (Ayotte et al., 2002; Cousineau et al., 2012). For comparison, we also presented the same clips with a local temporal change.

Participants

Twelve amusic (all amusics but A11) and 12 matched controls (10 of whom participated in Experiment 1A) were tested here.

Material

A subset of 24 (8 happy, 8 sad, and 8 peaceful) clips were selected. The clips intended to express fear were excluded because detection of errors was difficult in these clips, even for controls. Pitch-shift modified versions of the 24 original clips were created in which the melodic line of one entire measure was shifted either up or down by one semitone. Similarly, 18 original clips (6 happy, 6 sad, and 6 peaceful) were modified by changing the tone onset times randomly in one measure. The mean duration of the pitch-shift was 3.07 sec (range: 1.0-5.6) and the mean duration of the random time onsets was 3.25 sec (range: 1.24-6.06). The mean duration of the clips was 12.9 sec (range: 9.7-16.4) in the pitch-shift condition and 12.4 sec (range: 9.3-16.4) for the random time condition. Examples of both the pitch-shifted and random timing clips can be heard at www.brams.umontreal.ca/plab_ download.

Procedure

The type of modification was blocked and consisted of 48 trials in the pitch-shift condition and 36 trials in the random-time condition. The two conditions were presented to amusics and controls in counterbalanced order. The instruction was to monitor each excerpt so as to detect whether the pianist was “off-the-track or absent-minded” at some point during the performance. Participants responded “yes” if they detected an error and “no” otherwise. They were not informed of the nature of the changes, and no feedback was provided, with the exception of the practice examples (one example for each condition and for each emotion). The task lasted approximately 30 min.

Results and comments

We computed the nonparametric index of sensitivity A0 (Macmillan & Creelman, 1996) for each participant in each condition, by considering a “yes” response as a hit when there was a change and a “yes” response to an unchanged stimulus as a false alarm. Responses bias (B’’) was also computed for each individual. As expected, amusics performed poorly in the pitch-shift condition, and significantly worse than controls, $t(22) = 13.10$, $p < .001$ (Fig. 2). While most amusics performed within the normal range in the random-time condition, their scores still fell significantly below controls for the detection of time errors, $t(22) = 2.44$, $p < .05$. The responses bias measure (B’) indicated that amusics were significantly more conservative than controls in the pitch-shift condition [amusics' mean (SD): .22 (.22); controls' mean (SD): .46 (.25); $t(22) = 2.46$, $p < .05$], but not in the random-time condition [amusics' mean (SD): .14 (.41); controls' mean: .30 (.37); $t(22) = .99$, $p = .33$].

These results confirm that amusics' perceptual difficulties are not restricted to the melodies used for diagnosis. Performance on the pitch-shift condition was correlated with both the scores obtained on the MBEA scale test (Peretz et al., 2003), $r(22) = .93$, $p < .002$, and the scores obtained in the online outof-key test (Table 1; Peretz et al., 2008), $r(22) = .85$, $p < .001$. These correlations failed to reach significance when only amusics were considered, $r_s(10) = .44$ and .17, both $p = .16$. Nevertheless, the results show that the perceptual deficit that characterizes amusics extends to the processing of the musical clips from which amusics appear to be able to derive the emotional tone.

[Insert Figure 2 here]

This dissociation between perception (Experiment 1B) and emotion recognition (Experiment 1A) is consistent with the lack of correlations between the average score obtained in the pitch-shift and random-time condition and the emotional ratings, $r(19) = .11$, .38, .33, and .36, for happy, sad, scary and peaceful clips, respectively, $p > .05$. Note that the lack of correlation was also observed when the pitch-shift and randomtime condition were considered separately. Similar results were obtained when considering amusics only, $r_s(9)$ s .04, .05, .36, and .16, respectively; all $p > .05$.

Experiment 2: the contribution of mode and tempo

What has not yet been explained is how amusics can judge emotions in music, despite their difficulty in detecting obvious semitone pitch shifts in the same music. One hypothesis that we examine here experimentally is that amusics may rely more heavily on tempo (as suggested by Ayotte et al., 2002), since their pitch deficit might prevent them from using the musical mode, which critically depends on the semitone distinction, to distinguish major/happy from minor/sad music. The central objective of this experiment was to evaluate this hypothesis with a set of musical excerpts in which mode and tempo have been selectively manipulated (Peretz et al., 1998).

Participants

Eleven amusics (all amusics but A10 and A11) and 11 matched controls took part in this experiment.

Material

The 32 excerpts used in this experiment (previously employed in Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Peretz et al., 1998; Schellenberg, Peretz, & Vieillard, 2008) were selected from pre-existing Western classical music so that half of the excerpts evoked happiness and half evoked sadness. The excerpts were transcribed for piano and computer-generated using a piano timbre. The happy excerpts were in a major mode and played at a relatively fast tempo (between 80 and 255 BPM), while the sad excerpts were in a minor mode and played at a relatively slow tempo (between 20 and 100 BPM; for more details, see Peretz et al. 1998, and www.brams.umontreal.ca/plab_ download). Each original excerpt was manipulated electronically to create a mode inversion from major to minor and vice versa, using the same conventional procedure as Hevner (1935). Depending on the musical excerpt, this manipulation introduced variable frequency of semitones changes (from 2.7 to 46.3). After data collection, one sad excerpt was discarded from the analyses as its mode was ambiguous. The inclusion or exclusion of this excerpt did not affect the results. A neutralized tempo version was also created by adjusting all tempi to a unique value (with the quarter note = 84 MM) that corresponded to the median of all original tempi.

We performed the same acoustical analyses on the stimuli for this experiment as that done previously for the musical clips of Experiment 1A. The values are presented in Table 4 and showed that a number of acoustical characteristics differed between the original sad and happy stimuli as in Experiment 1. The mean pitch was higher, $t(29) = 5.27$, $p < .001$, the key was clearer, $t(29) = 2.70$, $p < .05$, roughness was greater, $t(29) = 3.64$, $p < .005$, and RMS energy was greater, $t(29) = 3.31$, $p < .005$, in happy than in sad stimuli. Pulse clarity, $t(29) = 1.89$, $p = .07$, and brightness, $t(29) = 1.02$, $p = .32$, did not differ significantly.

[Insert Table 4 here]

Procedure

The participants were presented with the original versions, as well as the mode and tempo change versions, of the same 32 excerpts. The presentation order of the excerpts was pseudorandomized so that the same excerpt was never presented twice in succession and that the same intended emotion (happy or sad) was not repeated more than 3 times. Four stimuli (2 sad, 2 happy), which were not used in the study, served as examples; no feedback was provided. The task was to judge to what extent the excerpts expressed sadness or happiness on a scale from 1 (sad) to 10 (happy). The stimuli were presented in a soundproof room through Genelec 8040A speakers. We also recorded the activity of the zygomatic and corrugator facial muscles and measured galvanic skin responses. Since these physiological measurements were too variable to be meaningful in both amusics and controls, the data will not be presented.

Results and comments

Amusics' mean ratings were similar to those of controls (Fig. 3A). The two-way mixed-design ANOVA computed on the mean raw ratings for the original version of the stimuli as a function of Group (amusic vs control) and intended Emotion (happy vs sad) confirmed that there was no reliable effect of Group, $F(1, 20) = 3.69$, $p = .07$, $\eta^2_{partial} = .16$, but a large effect of Emotion, $F(1, 20) = 671.06$, $p < .001$, $\eta^2_{partial} = .97$, and no interaction [$F(1, 20) = 1.97$, $p < .18$, $\eta^2_{partial} = .09$]. We also computed correlations between participants' pitch discrimination abilities (using the scores obtained in the pitch change detection task, Table 1) and the ratings provided for the original version of the stimuli. No relationships were found [for happy stimuli: $r(20) = .14$, $p = .52$; sad stimuli: $r(20) = .35$, $p = .12$], even in amusics when considered separately [$r(9) = .22$ and .14, respectively].

[Insert Figure 3 here]

In order to focus on the individual contributions of mode and tempo to the emotional judgments, we subtracted each rating given to the manipulated versions of an excerpt in the mode and tempo change conditions from the rating given to its original version. A mean score of 0 would indicate that there is no effect of the manipulation, while a negative score would indicate a change in ratings toward happiness, and a positive score a change toward sadness. As shown in Fig. 3B, amusics' ratings seemed to differ from controls' on the mode change version. This was supported by a three-way mixeddesign ANOVA computed on the mean subtracted ratings with Group (amusic vs control), intended Emotion (happy vs sad), and Structure (mode vs tempo) as factors. The ANOVA revealed a triple interaction, $F(1, 20) = 5.47$, $p < .05$, $\eta2_{partial} = .22$. Below, we report the results of follow-up ANOVAs performed on the tempo and mode change versions separately.

The groups showed similar sensitivity to the change of tempo, $F(1, 20) = 231.73$, $p < .001$, $\eta^2_{partial} = .92$. There was no Group effect ($F<1$) nor interaction between Group and intended Emotion (F<1). Thus, neutralization of tempo had the same effect on both amusics and controls' emotional ratings. In contrast, amusics did not respond to a change of mode as markedly as controls did, with a significant interaction between Group and intended Emotion, $F(1, 20) = 10.49$, $p < .005$, $\eta^2_{partial} = .34$. The group difference reached signifi- cance for the sad excerpts, $t(20) = 4.71$, $p < .001$, but not for the happy ones, $t(20) = 1.04$, $p = .31$. Nevertheless, amusics were sensitive to the mode manipulation; their ratings differed significantly from zero for both the happy and sad stimuli, $t(10) = 2.50$ and 5.63, respectively, $p < .05$, two-tailed tests. Although amusics were less sensitive to mode changes than controls were, there was no evidence that they relied more heavily on tempo at the group level. However, at the individual level, one can see that a few amusics (Happy: A4, A6 and A12; Sad: A4) were more affected (on the order of 2 SD) by the tempo neutralization (Fig. 3C).

In order to examine whether the reduced use of mode by amusics in emotional judgments compared to controls originated from their poor pitch discrimination, we examined the influence of the number of semitones changed by the mode manipulation in a given musical excerpt on its corresponding rating. We predicted that the more semitone changes in an excerpt, the more that would influence the ratings. Interestingly, only controls were found to be sensitive to the frequency of semitone changes, $r(29) = .38$, $p < .05$ (Fig. 4A). The frequency of semitone changes did not affect amusics' ratings, $r(29) = .00$, $p = .99$.

To examine what other features could have influenced amusics' judgments in the mode change condition, we compared the subtracted ratings with the values of the acoustical cues that were computed for the original versions minus their mode change versions. Of note, the mode change did not create any major differences in mean pitch. For example, the mean pitch of the original Vivaldi excerpt was 61.8, and it was 61.5 in the mode-change version.

Nevertheless, controls showed sensitivity to these small pitch differences, $r(29) = .67$, $p < .001$, whereas amusics did not, $r(29) = .27$, $p = .15$ (see Fig. 4B). Similarly, amusics' judgments did not show sensitivity to key clarity, $r(29) = .21$, $p = .26$, while controls' differences in rating did, $r(29) = .39$, $p < .05$ (Fig. 4C). In contrast, amusics were sensitive to timbral differences, as were controls. Differences in brightness, $r(29) = .38$, $p < .05$, in RMS energy, $r(29) = .43$, $p < .05$, and in roughness, $r(29) = .40$, $p < .05$, predicted the difference in ratings between the original and the mode change version in amusics. The correlations in controls for these characteristics were $r(29) = .50$, $p < .01$; .48, $p < .01$; and .51, $p < .005$, respectively. Neither amusics' nor controls' judgments showed sensitivity to pulse clarity, r(29) = .16, $p = .39$; $r(29) = .02$, $p = .93$, respectively.

Additionally, we performed correlations between the raw ratings obtained for the original stimuli and the acoustical characteristics. In this case, both amusics' and controls' ratings correlated with mean pitch [$r(29) = .73$ and .72, $p < .001$, respectively], RMS energy [$r(29) = .67$ and .61, $p < .001$], pulse clarity [$r(29) = .36$ and .37, $p < .05$], key clarity [$r(29) = .44$ and .42, $p < .05$] and roughness [$r(29) = .61$ and .61, $p < .001$], but not brightness [$r(29) = .18$ and .19, $p = .35$ and .31, respectively], as observed in Experiment 1.

[Insert Figure 4 here]

General Discussion

The results show that many amusic individuals have a remarkable sparing of emotional responses to music in the context of severe and lifelong deficits in processing music (see Table 1 for a summary of individual data). The finding of spared emotion recognition in a perceptually impaired system raises the fundamental question of the separability between emotion and cognition. This is not a new consideration. As mentioned in the introduction, a similar dissociation had been observed in I.R., a middle-aged woman without musical education, who suffered from irreversible deficits in music perception and memory as a consequence of a sudden brain accident (Peretz, Belleville, & Fontaine, 1997; Peretz & Gagnon, 1999). I.R. was found to be able to use the mode (major and minor) in which music was played in order to judge if its emotional tone was happy or sad, as normal participants did (Peretz et al., 1998). However, I.R. did not have the pitch deficit experienced by the amusics tested here, as I.R. was able to discriminate two successive tones a semitone apart (Peretz et al., 2001). Nevertheless, the present findings extend the dissociation between emotions and perception in music to congenital amusia and raise a set of novel questions regarding its origin.

The dissociation between the perceptual and emotional processing of music is important for theoretical models. For example, we (Peretz & Coltheart, 2003) have posited a common perceptual system for both music identity recognition and emotion expression recognition. In other domains, such as facial processing, it has been proposed that perceptual processing is distinct and is implemented in a separate functional (and neural) system for identity than that used for emotion processing of faces (Bruce & Young, 1986; Haxby, Petit, Ungerleider, & Courtney, 2000). According to the classical view, prosopagnosic individuals suffer from deficits in facial identity processing with no difficulties in processing other aspects of faces, particularly facial expressions (Duchaine, Parker, & Nakayama, 2003; Humphreys, Avidan, & Behrmann, 2007).

In the music domain, the task characteristics may play a more critical role than in facial processing. Asking participants whether music evokes some basic emotion makes sense to nonmusicians because this is one of the primary reasons why most people listen to music. Asking nonmusicians to detect an error in music is much less common and is probably never done intentionally. Thus, by being closer to everyday listening experiences, emotional judgments may be more appropriate to reveal the content and the organization of the listeners' implicit knowledge of music structure than most non-emotional judgments (Peretz et al., 1998).

Emotional judgments rely on a large variety of structural cues. The most potent musical cues, also the most frequently studied (Gabrielsson & Lindström, 2010), are mode, tempo, dynamics, articulation and timbre. Here, we examined the role of mode and tempo experimentally (Experiment 2) and assessed the contribution of correlated attributes, such as mean pitch, key and pulse clarity, and timbral cues (brightness, energy, and roughness) in both Experiment 1A and 2. Amusics' judgments were mostly influenced by temporal (tempo, pulse clarity) characteristics and timbre. There was no evidence that amusics could use semitone differences. In contrast, controls' emotional judgments exhibited sensitivity to these features. This difference in access to subtle emotional cues may explain why amusics were impaired but still above chance in judging the happy-sad character of the music when the mode was experimentally changed. The deficit was expected as a result of their impaired pitch perception system, which affects their ability to detect the critical semitone difference that differentiates the minor from the major mode (e.g., Hyde & Peretz, 2004) and the pitch salience (harmonicity) that differentiates consonant from dissonant chords (Cousineau et al., 2012). Our novel finding here is that amusics are able to use other timbre-related features to judge the emotion conveyed by music. In sum, a musical emotion processing deficit is present but subtle in congenital amusia, and can be compensated for by the use of tempo, pulse clarity, large mean pitch differences, and timbre in most circumstances.

One implication of the present findings is that most amusic individuals should enjoy listening to music. Yet, as noted in the introduction, amusics vary widely in their interest in music (Ayotte et al., 2002; McDonald & Stewart, 2008; Omigie et al., 2012) and this variability appears unrelated to the severity of the disorder (McDonald & Stewart, 2008). Here, we replicate this observation and support it in a quantified manner, as evident through close examination of the individual data (Table 1). For example, while a relationship seems to exist in participant A6 between the severity of the musical pitch deficit, the emotion recognition performance, and musical appeal, there is no clear relationship in participant A8. A6 seems impaired in all music-related activities and dislikes music. In contrast, participant A8, who is a severe case of amusia, exhibits normal recognition of emotions and responds to music as controls do according to his questionnaire responses. Such a dissociation pattern is consistent with the recent discovery of music anhedonia, a deficit that can occur in isolation in 20% of the normal population (Mas-Herrero et al., 2014). Thus, perception, emotion and enjoyment are possibly distinct components in music processing. Understanding their inter-relationship remains a major challenge for future studies in neuroscience of music.

An interesting issue raised by the present findings concerns domain specificity. From a music-theoretic perspective, the perception of a mode change relies on the detection of a scale degree difference, hence on knowledge of the rules of the Western tonal system. Such knowledge would be both culture- and music-specific. Accordingly, a fine-grained pitch deficit should affect music selectively. However, recent findings question the domain specificity of the present findings. First, a mild deficit in identifying happiness and sadness from speech prosody has been observed in congenital amusia (Thompson, Marin, & Stewart, 2012). Second, the spectra of major intervals have been found to be more similar to the spectra of excited speech, whereas the spectra of minor intervals were more similar to the spectra of subdued speech (Bowling, Gill, Choi, Prinz, & Purves, 2010; Daniel Liu Bowling et al., 2012; Curtis & Bharucha, 2010). These similarities may reflect a common evolutionary origin for music and vocal emotions (Spencer, 1857) or a musical neural invasion ("neural recycling", to adopt the terminology of Dehaene & Cohen, 2007). Musical emotions might recruit the same circuits that have evolved for vocal emotions (e.g. Peretz, 2010). Further comparative study of music and prosody in amusia may shed further light on this issue.

We can conclude from the present study that severe diffi- culties in music discrimination and memory can spare emotional processing of music to a large extent. This preservation may serve as a springboard for intervention, whereas systematic research on the residual difficulties may open new avenues for testing theories regarding the biological origins of musical emotions.

Acknowledgements

The authors thank the participants for their continued collaboration, Mihaela Felezeu for assistance in testing participants, Christine Becket for the musical analyses, Dawn Merrett for editing the manuscript. The work was supported by grants from the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC) and a Canada Research Chair in neurocognition of music to I.P., by a NSERC-Create graduate fellowship in Auditory Cognitive Neuroscience and a graduate scholarship from the CIHR to S.P.

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., … Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*, *136*(5).

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain : A Journal of Neurology*, *125*(Pt 2), 238–51. http://doi.org/10.1093/brain/awf028

Bowling, D. L., Gill, K., Choi, J. D., Prinz, J., & Purves, D. (2010). Major and minor music compared to excited and subdued speech. *The Journal of the Acoustical Society of America*, *127*(1), 491. http://doi.org/10.1121/1.3268504

Bowling, D. L., Sundararajan, J., Han, S., Purves, D., Touma, H., Patel, A., … Jürgens, U. (2012). Expression of Emotion in Eastern and Western Music Mirrors Vocalization. *PLoS ONE*, *7*(3), e31942. http://doi.org/10.1371/journal.pone.0031942

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. http://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Cousineau, M., McDermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19858–63. http://doi.org/10.1073/pnas.1207989109

Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*(3), 335–348. http://doi.org/10.1037/a0017928

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), 1–10. http://doi.org/10.1016/S0010-0277(00)00136-0

Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. http://doi.org/10.1016/j.neuron.2007.10.004

Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, *32*(7), 827–38. http://doi.org/10.1068/P5067

Eerola, T., & Toiviainen, P. (2004). MIDI toolbox: MATLAB tools for music research.

Gabrielsson, A., & Lindström, E. (2010). *The role of structure in the musical expression of emotions*. *Handbook of music and emotion: Theory, research, applications*. new York, NY: Oxford University Press.

Gosselin, N., Peretz, I., Hasboun, D., Baulac, M., & Samson, S. (2011). Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. *Cortex*, *47*(9), 1116–1125. http://doi.org/10.1016/j.cortex.2011.05.012

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244. http://doi.org/10.1016/j.neuropsychologia.2006.07.012

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., & Samson, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, *128*(3).

Haxby, J. V., Petit, L., Ungerleider, L. G., & Courtney, S. M. (2000). Distinguishing the Functional Roles of Multiple Regions in Distributed Neural Systems for Visual Working Memory. *NeuroImage*, *11*(5), 380–391. http://doi.org/10.1006/nimg.2000.0592

Hevner, K. (1935). The Affective Character of the Major and Minor Modes in Music. *The American Journal of Psychology*, *47*(1), 103. http://doi.org/10.2307/1416710

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research*, *176*(2), 356–373. http://doi.org/10.1007/s00221-006-0621-5

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical Thickness in Congenital Amusia: When Less Is Better Than More. *Journal of Neuroscience*, *27*(47), 13028–13032.

Hyde, K. L., & Peretz, I. (2004). Brains That Are out of Tune but in Time. *Psychological Science*, *15*(5), 356–360. http://doi.org/10.1111/j.0956-7976.2004.00683.x

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., Peretz, I., Ashburner, J., … Evans, A.

(2006). Morphometry of the amusic brain: a two-site study. *Brain : A Journal of Neurology*, *129*(10), 2562–70. http://doi.org/10.1093/brain/awl204

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI Evidence of an Abnormal Neural Network for Pitch Processing in Congenital Amusia. *Cerebral Cortex February*, *21*, 292–299. http://doi.org/10.1093/cercor/bhq094

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1812. http://doi.org/10.1037/0096-1523.26.6.1797

Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*, 1–8.

Lebrun, M.-A., Moreau, P., McNally-Gagnon, A., Mignault Goulet, G., & Peretz, I. (2012). Congenital amusia in childhood: A case study. *Cortex*, *48*(6), 683–688. http://doi.org/10.1016/j.cortex.2011.02.018

Lima, C. F., & Castro, S. L. (2011). Emotion recognition in music changes across the adult life span. *Cognition & Emotion*, *25*(4), 585–598. http://doi.org/10.1080/02699931.2010.502449

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone Deafness: A New Disconnection Syndrome? *Journal of Neuroscience*, *29*(33).

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of “nonparametric” measures of sensitivity and response bias. *Psychonomic Bulletin & Review*, *3*(2), 164–170. http://doi.org/10.3758/BF03212415

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual Differences in Music Reward Experiences. *Music Perception: An Interdisciplinary Journal*, *31*(2), 118–138.

Mas-Herrero, E., Zatorre, R., & Rodriguez-Fornells, A. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology*, *24*(6), 699–704.

McDonald, C., & Stewart, L. (2008). Uses and functions of music in congenital amusia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 345–355. http://doi.org/10.1525/mp.2008.25.4.345

Mignault Goulet, G., Moreau, P., Robitaille, N., Peretz, I., Sarkamo, T., Tervaniemi, M., … Gabrieli, J. (2012). Congenital Amusia Persists in the Developing Brain after Daily Music Listening. *PLoS ONE*, *7*(5), e36860. http://doi.org/10.1371/journal.pone.0036860

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition*, *81*(3), 337–344. http://doi.org/10.1016/j.bandc.2013.01.004

Omigie, D., Dellacherie, D., Hasboun, D., George, N., Clement, S., Baulac, M., … Samson, S. (2015). An Intracranial EEG Study of the Neural Dynamics of Musical Valence Processing. *Cerebral Cortex (New York, N.Y. : 1991)*, *25*(11), 4038–47.

Peretz, I. (2010). Towards a Neurobiology of Musical Emotions. In J. Juslin, P; Sloboda (Ed.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 99–126). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199230143.003.0005

Peretz, I. (2013). The biological foundations of music: insights from congenital amusia. *The Psychology of Music*, (3), 551–564.

Peretz, I., Belleville, S., & Fontaine, S. (1997). Dissociations entre musique et langage après atteinte cérébrale: Un nouveau cas d'amusie sans aphasie. *Canadian Journal of Experimental Psychology*, *51*(4), 354–368. http://doi.org/10.1037/1196-1961.51.4.354

Peretz, I., Blood, A. J., Penhune, V., & Zatorre, R. (2001). Cortical deafness to dissonance. *Brain*, *124*(5), 928–940.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain*, *132*(5), 1277–1286.

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology*, *58*(3), 478–482. http://doi.org/10.1002/ana.20606

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of Musical Disorders. *Annals of the*

New York Academy of Sciences, *999*(1), 58–75. http://doi.org/10.1196/annals.1284.006

Peretz, I., & Coltheart, M. (2003). Modularity of music processing, *6*(7), 688–691. http://doi.org/10.1038/nn1083

Peretz, I., Cummings, S., & Dubé, M.-P. (2007). The Genetics of Congenital Amusia (Tone Deafness): A Family-Aggregation Study. *The American Journal of Human Genetics*, *81*(3), 582–588. http://doi.org/10.1086/521337

Peretz, I., & Gagnon, L. (1999). Dissociation between recognition and emotional judgements for melodies. *Neurocase*, *5*(1), 21–30. http://doi.org/10.1080/13554799908404061

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. http://doi.org/10.1016/S0010-0277(98)00043-2

Quarto, T., Blasi, G., Pallesen, K. J., Bertolino, A., & Brattico, E. (2014). Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits. *PloS One*, *9*(7), e103278. http://doi.org/10.1371/journal.pone.0103278

Schellenberg, E. G., Peretz, I., & Vieillard, S. (2008). Liking for happy- and sad-sounding music: Effects of exposure. *Cognition & Emotion*, *22*(2), 218–237. http://doi.org/10.1080/02699930701350753

Sloboda, J. A., Wise, K., & Peretz, I. (2005). Quantifying Tone Deafness in the General Population. *Annals of the New York Academy of Sciences*, *1060*(1), 255–261.

Spencer, H. (1857). The origin and function of music. *Essays, Scientific, Political, and Speculative*, *2*.

Thompson, W. F. W., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *PNAS*, *109*(46), 19027–19032. http://doi.org/10.1073/pnas.1210344109

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752. http://doi.org/10.1080/02699930701503567

Vuvan, D. T., Nunes-Silva, M., & Peretz, I. (2015). Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia. *Cortex*, *69*, 186–200. http://doi.org/10.1016/j.cortex.2015.05.002

Zendel, B. R., Lagrois, M.-É., Robitaille, N., & Peretz, I. (2015). Attending to Pitch Information Inhibits Processing of Pitch Information: The Curious Case of Amusia. *Journal of Neuroscience*, *35*(9).

Article supplémentaire: Tableaux et Figures

Tableau I. A Supp.: Caractéristiques des participants

Characteristics of participants and individual scores obtained by amusics on the six tests (global score) and on the scale, contour and interval tests (melodic composite score) on the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003). The average scores obtained on the online out-of-key test (Peretz et al., 2008), the pitch change detection task (Hyde & Peretz, 2004), and the Affective Response to Music questionnaire (McDonald & Stewart, 2008) are also provided. Preserved (+; within 2 SD of normal scores) or altered (-; 2 SD away from normal mean) performance observed in Experiment 1 and 2 are also summarized. Amusic and control participants' characteristics and scores are expressed with the mean (SD).

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	Amusics	Controls
Characteristics															
Gender	F	F	F	F	F	F	M	M	M	F	M	F	M	8F, 5M	10F, 5M
Age (years)	62	68	68	66	62	69	62	65	66	57	54	69	67	64.2	62.9 (3.8)
Education (years)	18	15	8	14	20	15	19	19	14	19	15	21	19	16.6	15.6 (2.6)
Musical training#	1	2	0	2	1	0	1	1	1	3	1	4	1	1.4	1.6 (1.6)
MBEA															
Global score (/30)	22.2	20.8	18.8	18.8	21.8	18	19	15.3	18	19	16	22.3	17	19	27.2 (1.3)
Melodic composite score (/30)	22.7	20.7	15.7	19.3	19.7	17	17.7	15	17.7	16	16.3	20	16.3	18	27.2 (1.7)
Pitch change detection∞ (% hits - false alarms)	80	65.7	51.5	69.3	57.1	35.6	75.2	50.9	72.2	63.9	n.a.	70.4	87.4	64.9	96.7 (2.6)
Online out-of-key test (% correct)	63	29.2	46	45.8	62.5	54	50	20.8	37.5	50	45.8	79.2	45.8	48.4	90.1 (6.4)
Affective questionnaire (/5)	2.4	1.4	1.9	1.8	1.8	4.2	3.0	2.5	1.3	2.7	n.a.	2	1.2	2.2	1.7 (0.6)
Experiment 1															
Emotion recognition	+	−sc	+	+	+	−sc, pe	+	+	+	+	−pe	+	n.a.		
Pitch shift detection (A')	−	−	−	−	−	−	−	−	−	−	n.a.	−	−		
Time error detection (A')	+	−	+	−	+	−	+	−	+	+	n.a.	+	+		
Experiment 2															
Raw ratings original condition	−ha	+	+	+	+	+	+	+	+	n.a	n.a	+	+		
Mode change condition	+	+	+	+	+	+	+	−sa, ha	−sa	n.a.	n.a.	−sa	−sa		

Note: F = female ; M = male; #Musical training is classified with level 1 = < one year; 2 = 1-3 years; 3 = 4-6 years; and 4 = 7-10 years. ∞average over ± 1/4, 1/2, 1 semitone changes; sc = scary; pe = peaceful; ha = happy; sa = sad; + = Preserved; − = Altered

Tableau II. A Supp.: Caractéristiques acoustiques des extraits présentés (étude 1a)

Mean pitch, or F0, and pitch range (in parentheses), expressed in MIDI code, with pitch names, and in Hz, as well as acoustic values extracted with the MIR Toolbox for the musical clips of Experiment 1, as a function of emotion.

	Mean pitch			Acoustic Features				
Emotion	MIDI	Label	Hz	Key clarity	Pulse clarity	Brightness	RMS energy	Roughness[N]
Happy	65.9 (60.5–71.3)	F#5 (C#5-B5)	431.9 (302.5–589.7)	.82 (.74–.96)	.80 (.65–.88)	.25 (.21–.28)	.17 (.12–.21)	.09 (.06–.13)
Sad	59.3 (55.1–65.7)	B4 (G4-F#5)	299.4 (234.2–464.7)	.80 (.58–.92)	.80 (.65–.89)	.25 (.23–.28)	.14 (.12–.17)	.05 (.02–.10)
Scary	56.1 (41.3–74.1)	G#4 (F3-D6)	318.3 (94.7–718.8)	.68 (.45–.87)	.56 (.30–.87)	.24 (.19–.33)	.13 (.07–.19)	.07 (.01–.13)
Peaceful	62.1 (55.4–66.9)	D5 (G4-G5)	388.4 (252.8–574.7)	.80 (.66–.89)	.83 (.77–.89)	.24 (.22–.26)	.15 (.10–.18)	.07 (.02–.09)

[N] normalized with minimum and maximum roughness values of the dataset.

Tableau III. A Supp.: Jugements émotionnels des amusique et contrôles (étude 1a)

Recognition by amusics and matched controls of happiness, sadness, scariness (fear) and peacefulness evoked by music, demonstrated with mean ratings (on a scale of 0–9) of how much each clip expressed each of the four intended emotions. Means (and standard errors).

	Perceived emotions			
Intended emotions	Happiness	Sadness	Scary	Peacefulness
Amusics				
Happiness	7.8 (.2)	.1	.1	2.6
Sadness	1	5.6 (.4)	.8	2.8
Fear	1.7	2.4	4.2 (.6)	.7
Peacefulness	3.2	2.8	.1	4.9 (.5)
Controls				
Happiness	7.8 (.3)	.2	.1	1.9
Sadness	.8	6.1 (.5)	.6	4.4
Fear	1.2	3.1	5.2 (.4)	.5
Peacefulness	3.3	3.1	.2	6.2 (.4)

Tableau IV. A Supp.: Caractéristiques acoustiques des extraits présentés (étude 2)

Mean pitch (F0), as expressed in MIDI code, with pitch names and in Hz, as well as the acoustic values extracted with the MIR Toolbox for the happy and sad music of Experiment 2. Range in parentheses.

	Mean Pitch			Acoustic Features				
Emotion	MIDI	Label	Hz	Key Clarity	Pulse Clarity	Brightness	RMS energy	Roughness[N]
Happy	66.2 (61.4-71.4)	F#5 (C#5-B5)	473.4 (343.2-652.8)	.85 (.67-.96)	.79 (.63-.90)	.15 (.08-.20)	.17 (.13-.22)	.11 (.05-.22)
Sad	59.6 (50.9-68.4)	C5 (D#4-G#5)	315.8 (181.3-524.07)	.75 (.52-.93)	.69 (.16-.91)	.16 (.14-.20)	.14 (.11-.21)	.07 (.04-.14)

[N] normalized with minimum and maximum roughness values of the dataset

Figure 1. A Supp.: Jugements émotionnels des amusique et contrôles (étude 1a)

Individual average rating as a function of group and intended emotion in Experiment 1A. Mean group values are indicated by columns.

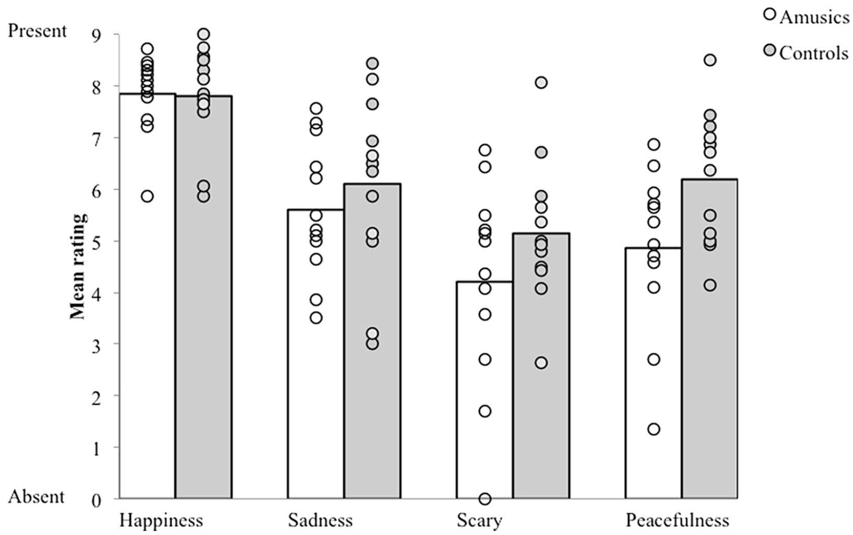


Figure 2. A Supp.: Changements de hauteur tonale vs temporels (étude 1b)

Individual mean A′ obtained in the detection of a pitch-shift as a function of the detection of a random time change in Experiment 1B.

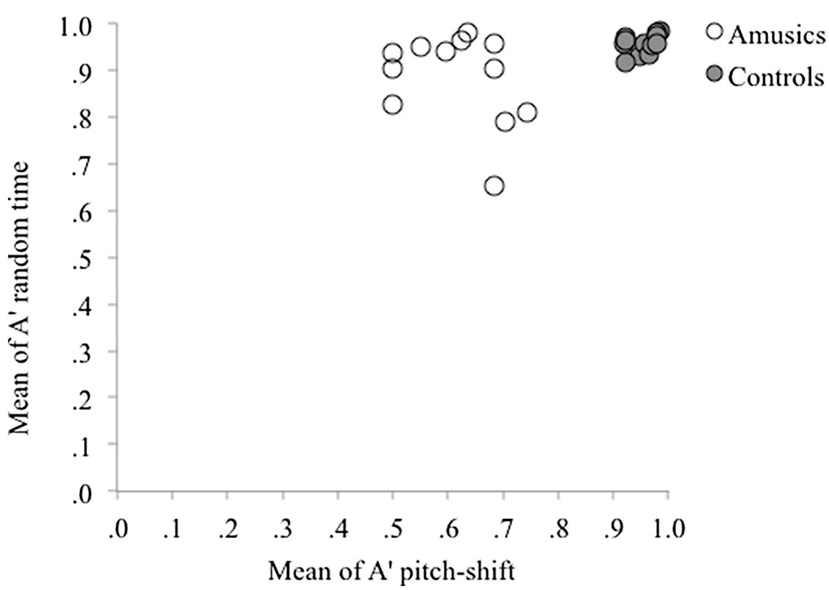


Figure 3. A Supp.: Jugements émotionnels des amusique et contrôles (étude 2)

Individual average rating as a function of group and intended emotion in Experiment 2 for the original stimuli (A). Difference ratings for the original stimulus minus the manipulated stimulus are presented for the inverted mode in (B) and the neutral tempo in (C). Mean group values are indicated by columns.

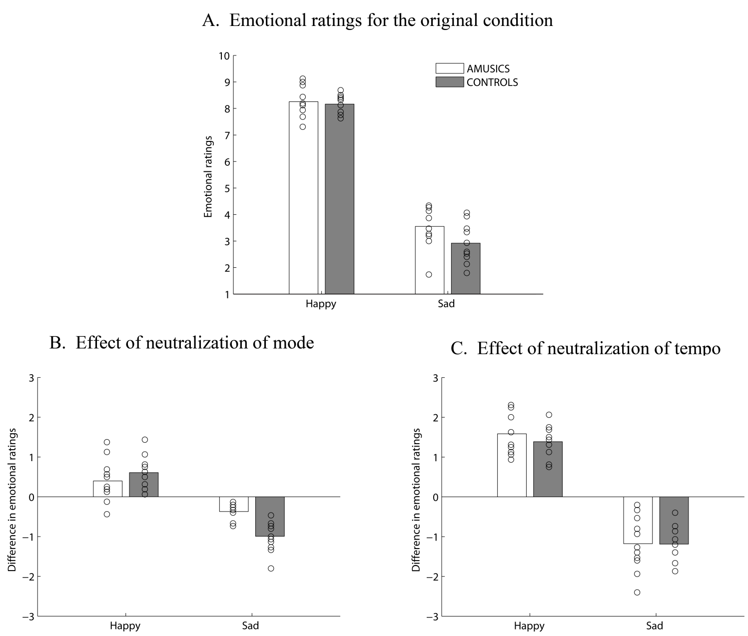
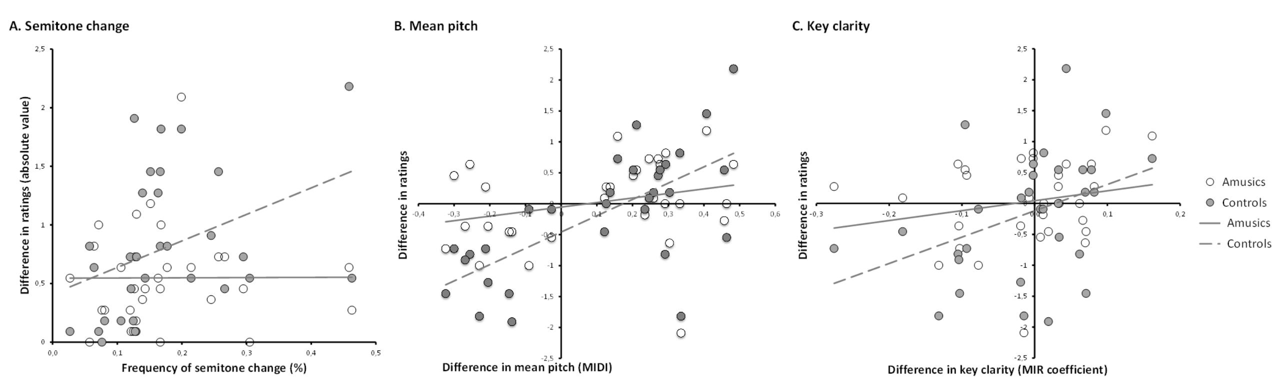


Figure 4. A Supp.: Corrélations avec les caractéristiques acoustiques (étude 2)

Individual average difference ratings for the original minus the inverted mode stimulus as a function of the frequency of semitone changes (A), the difference in mean pitch (B) and in key clarity (C) for amusics and controls in Experiment 2. Note that negative differences were obtained when the acoustical value was higher in the mode change condition than in the original version. A mean score of 0 indicates no effect of the manipulation.



Article Supplémentaire: Matériel Supplémentaire

Affective Responses to Music questionnaire

Part two of the *Uses and Function questionnaire* (McDonald & Stewart, 2008).

1. A certain song can often evoke nostalgic memories of past times or a past event for me.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
1	2	3	4	5

2. I have never experienced tingles/goose pimples/shivers from any kind of music.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
5	4	3	2	1

3. Certain music can put me in a good mood or ‘match’ my current mood.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
1	2	3	4	5

4. I have never felt that music has moved me to tears/catharsis/ or emotional release.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
5	4	3	2	1

5. Music can calm/ soothe/ relax or relieve my stress at times.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
1	2	3	4	5

6. I rarely experience the feeling of music uplifting my mood.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
5	4	3	2	1

7. Occasionally, certain music can sadden my mood.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
1	2	3	4	5

8. I don’t find that music can comfort me in uncertain times.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
5	4	3	2	1

9. Certain music can sometimes motivate or excite me.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
1	2	3	4	5

10. I don’t associate music with the onset of spiritual feeling/ experience.

Strongly agree	Agree	Unsure	Disagree	Strongly disagree
5	4	3	2	1

Annexe 2. Autres projets réalisés durant le doctorat

Lehmann, A., & Paquette, S. (2015). Cross-domain processing of musical and vocal emotions in cochlear implant users. *Frontiers in Neuroscience*, *9*.

Paquette, S., & Mignault Goulet, G. (2014). Lifetime benefits of musical training. *Frontiers in Neuroscience*, *8*, 89.

Paquette, S., Mignault Goulet, G., & Rothermich, K. (2013). Prediction, attention, and unconscious processing in hierarchical auditory perception. Frontiers in Psychology, *4*, 955.

Vuvan, D., Paquette, S., Mignault Goulet, G., Royal, I., Felezeu, M., & Peretz, I. (2017). The Montreal protocol for identification of amusia. *Behavior Research Methods*.

Agust, T. R., Paquette, S., Suied, C., Pressnitzer, D., & Belin, P. (sous Rev.). Voice selectivity in the temporal voice area without low-level acoustic confounds. *Scientific Reports*.

Paquette, S., Fournier, P., Dupon, S., Szabo, F., Galan, P., & Samson, S. (En Prep.). Prevalence of tinnitus in patients with surgical resection including the amygdala. *JAMA*.

Paquette, S., Lagrois, M-E., Peretz I. (En Prep.). Musik: A cross-sectional study on emotion in music composition in the general population. *Psychology of Music.*